AUSTRALIAN JOURNAL
OF
SCIENTIFIC RESEARCH

SERIES A
PHYSICAL SCIENCES

VOLUME 4

MELBOURNE
1951

AUSTRALIAN JOURNAL OF SCIENTIFIC RESEARCH

Published by the Council for Scientific and Industrial Research Organization and the Australian National Research Council

Issued in two Series :

Series A – Physical Sciences

(Four issues each year appearing March, June, September, December)

Series B – Biological Sciences

(Four issues each year appearing February, May, August, November)

Price : Each series 30/– per annum, separate issues 7/6 each

EDITORIAL BOARD

Chairman and Editor : N. S. Noble
Members : Professor Sir Macfarlane Burnet
Professor E. J. Hartung
Professor L. H. Martin
Professor J. G. Wood

All enquiries and manuscripts should be forwarded to :

The Editor,
Australian Journal of Scientific Research,
Commonwealth Scientific and Industrial Research Organization,
314 Albert Street, East Melbourne, C.2, Victoria.

Printed by Australasian Medical Publishing Co. Ltd., Sydney.

CONTENTS

Number 1, March 1951

PAGE

The Design of Reflecting Microscope Objectives. By W. H. Steel .. 1

Solution of a Heat Flow Problem. By J. R. M. Radok 12

The Momentum Spectrum and Positive Excess of Cosmic Ray Mesons at Sea-Level. By D. E. Caro, J. K. Parry, and H. D. Rathgeber .. 16

Observations of the Spectrum of High-Intensity Solar Radiation at Metre Wavelengths. IV. Enhanced Radiation. By J. P. Wild .. 36

Radio Observations of Two Large Solar Disturbances. By W. N. Christiansen, J. V. Hindman, A. G. Little, Ruby Payne-Scott, D. E. Yabsley, and C. W. Allen 51

Thermoelectric Properties of Titanium with Special Reference to the Allotropic Transformation. By H. W. Worner 62

Aqueous Diffusion Measurements by the Lamm Scale Method and Their Correlation with Membrane Diffusion Measurements. By D. G. Gillam .. H.. 84

Sulphonamides. II. Structure and Tautomerism of Sulphapyridine, Sulphathiazole, and Sulphanilylbenzamidine. By S. J. Angyal and W. K. Warburton 93

The Chemistry of Western Australian Plants. IV. *Duboisia hopwoodii.* By W. Bottomley and D. E. White 107

Short Communication

The Chemistry of Western Australian Plants. V. Vellein from *Velleia discophora.* By W. Bottomley and D. E. White 112

NUMBER 2, JUNE 1951

PAGE

Transmission-Like Multiple Beam Reflection Interferometry. By C. F. Bruce 117

Solar Radio-Frequency Emission from Localized Regions at Very High Temperatures. By J. H. Piddington and H. C. Minnett .. 131

Observations of the Source of Radio-Frequency Radiation in the Constellation of Cygnus. By B. Y. Mills and A. B. Thomas .. 158

The Influence of Slit-Width on the Shape and Intensity of Infra-Red Absorption Bands. By J. B. Willis 172

Rupture of Oxide Films during Repeated Sliding. By A. J. W. Moore and W. J. McG. Tegart 181

Studies of the Wallach Transformation. II. Steric Effects in Some Symmetrically Disubstituted Azoxybenzenes. By P. H. Gore and G. K. Hughes 185

Liquid-Vapour Equilibria. II. The System Benzene-*n*-Heptane. By I. Brown and A. H. Ewald 198

The Dielectric Properties of Symmetrical Long-Chain Secondary Alcohols in the Solid State. By R. J. Meakins and R. A. Sack .. 213

Corrigenda 229

Number 3, September 1951

PAGE

The Solution of Polynomial Equations with the Aid of the Electrolytic Tank. By L. Tasny-Tschiassny and A. G. Doe 231

Galactic Radiation at 18·3 Mc/s. By C. A. Shain 258

Angular Distribution of α-Particles and Protons from Li^6+D^2. By D. N. F. Dunbar and F. Hirst 268

Further Measurements of the Angular Distribution of α-Particles from the $Li^7(p,\alpha)He^4$ Reaction. By F. Hirst and R. Uebergang 284

The Surface Loss Correction for Nuclear Emulsions. By S. L. Martin .. 296

The Temperature Dependence of the Thermal Conductivity of Air. By W. G. Kannuluik and E. H. Carman 305

A Survey of the Stress between the Ocean and Atmosphere. By C. H. B. Priestley 315

The Structure of Steady-State Anticyclones. By R. W. James 329

Conjugated Compounds. III. Spectra and Ionization in Polyenes and Polymethines by a Finite Difference Method. By N. S. Bayliss and J. C. Rivière 344

The Dielectric Properties of a Series of *dl-n*-Eicosanols. By R. J. Meakins and H. K. Welsh 359

The Dielectric Properties of Crystalline Long-Chain *n*-Primary Alcohols at Low Frequencies. By R. J. Meakins and Joan W. Mulley .. 365

Some Electrophoretic Studies of Proteins in the Adsorbed State. By C. W. N. Cumper and A. E. Alexander 372

Adsorption and Hydrogenation of Gases on Transition Metals. By M. E. Winfield 385

The Crystal Structure of Potassium Hexafluoroplatinate (IV). By D. P. Mellor and N. C. Stephenson 406

The Crystalline Structure of Cellulose as Revealed by Acid Hydrolysis. By D. H. Foster and A. B. Wardrop 412

Experiments on the Synthesis of the Acridone Alkaloids. By G. K. Hughes and E. Ritchie 423

Short Communications

A Circuit for the Operation of a Synchrotron in conjunction with a Wilson Cloud Chamber. By E. G. Muirhead and H. Lichtblau 432

Small Scale Preparation of Methyl Red Indicator. By R. G. Gillis .. 435

Corrigenda 436

NUMBER 4, DECEMBER 1951

PAGE

Collision Processes Involving Highly Ionized Atoms. By E. R. Hill .. 437

The Straggling of Electrons below the Critical Energy. By K. C. Hines 450

Observations of Galactic Radiation at Frequencies of 1200 and 3000 Mc/s. By J. H. Piddington and H. C. Minnett 459

Galactic Radiation at Radio Frequencies. IV. The Distribution of Radio Stars in the Galaxy. By J. G. Bolton and K. C. Westfold 476

The Position and Movement on the Solar Disk of Sources of Radiation at a Frequency of 97 Mc/s. I. Equipment. By A. G. Little and Ruby Payne-Scott 489

The Position and Movement on the Solar Disk of Sources of Radiation at a Frequency of 97 Mc/s. II. Noise Storms. By Ruby Payne-Scott and A. G. Little 508

A High Speed Chronograph and its Application to the Time of Flight of Cosmic Rays. By V. C. Officer 526

Operational Impedance Matrices of *n*-Phase Partially Symmetrical Machines. By D. W. C. Shen 544

An Analysis of the Galvanometer Amplifier and its Response to Alternating Electromotive Forces and Mechanical Vibrations. By R. G. Wylie and A. F. A. Harper 560

The Photovoltaic Effect in Natural Lead Sulphide. By R. Lawrance.. 569

The Infra-Red Spectra of Porphyrins and their Iron Complexes. By J. E. Falk and J. B. Willis 579

Polarography with Alternating Currents. III. A.C. Polarography of Copper, Bismuth, Antimony, and Tin with Special Reference to the Presence of Halogen Ions. By B. Breyer, F. Gutmann, and S. Hacobian 595

Polarography with Alternating Currents. IV. A.C. Polarography of Uranium, Iron, Nickel, and Manganese. By B. Breyer and S. Hacobian 604

Polarography with Alternating Currents. V. A.C. Polarography of Some Common Anions. By B. Breyer and S. Hacobian 610

Dielectric Absorption in Solid Aliphatic Methyl Esters. By J. S. Dryden and H. K. Welsh 616

The Adsorption of Sodium Cetyl Sulphate on Cassiterite. By G. R. Edwards and W. E. Ewers 627

The Thermal Decomposition of Thorium Oxalate. By Roberta Beckett and M. E. Winfield 644

Coordination Complexes of Indium and 2,2′,2″-Terpyridyl and of Indium and 5-Nitro-1,10-Phenanthroline. By G. J. Sutton.. 651

Studies in the Chemistry of Some Thallium III Complexes. By G. J. Sutton 654

Corrigendum 658

Index to Volume 4 659

THE DESIGN OF REFLECTING MICROSCOPE OBJECTIVES

By W. H. Steel*

[*Manuscript received September* 26, 1950]

Summary

The general theory is developed and the exact solutions obtained for the design of reflecting microscope objectives having two or three concentric spherical surfaces. It is shown that the only serious optical defect of reflecting objectives, low contrast due to the high obstruction ratio, can be satisfactorily controlled in three-mirror systems.

Non-concentric objectives are discussed, but in practice prove to be difficult to adjust.

I. Introduction

In recent years, reflecting objectives have had many applications in ultra-violet or infra-red microscopy, possessing the obvious advantage over a refracting system of complete freedom from chromatic aberrations. These systems can be considered as falling into two classes : those based on a single concave mirror, and those using a concave and a convex reflector.

An objective using a concave reflector alone has the advantage of negative field curvature which can compensate the positive curvatures of an eyepiece. However, a single reflector cannot be corrected for both spherical aberration and coma. A spherical mirror with a stop at its centre is free from coma but the spherical aberration is only tolerable at numerical apertures below 0·18 (Bouwers 1946). At higher apertures the spherical aberration must be corrected by using refracting elements ; Johnson (1949) used a Mangin mirror in combination with lenses, while Bouwers (1946) used a meniscus lens. Both these systems must be achromatized by using two different refracting materials.

The Schwarzschild (1905) system of a convex and a concave reflector has been investigated by Burch (1947) who gave the general conditions for aplanatism. He showed that these gave one solution with spherical surfaces, but as in this case the obscuration was rather high, he preferred systems where one or both reflectors were aspherical. However, the objective using spherical reflectors has been used successfully by Brumberg (1941, 1943) and by Seeds and Wilkins (1949). Finally Grey and Lee (1949) and Grey (1949, 1950) have designed catadioptric objectives combining a two mirror system with refracting elements.

The Schwarzschild objective with spherical surfaces combines the two great practical advantages of ease of construction with complete freedom from chromatic aberrations, as no refracting elements are needed. In the present paper the method of designing these objectives is worked out and it is shown that the radii required for the surfaces are given by a simple series of equations. As

---

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

mentioned above, the convex mirror obscures a considerable fraction of the aperture and two methods of decreasing this obscuration are discussed : (i) the use of a non-concentric system, or (ii) the addition of a further concave reflector.

II. Monocentric Optical Systems

Both the single mirror and meniscus of Bouwers (1946) and the Schwarzschild objective with spherical surfaces are monocentric optical systems, all the surfaces having a common centre. When the stop is at this centre, all principal rays are axes of the system and no astigmatism or oblique aberrations are present. The only aberrations requiring correction are spherical aberration and, if some of the surfaces are refracting, longitudinal chromatic aberration. The field is curved, the ideal object and image surfaces being spherical with the same centre as the surfaces of the objective.

Unlike most other systems, the image distance at any aperture and hence the spherical aberration of a monocentric system can be expressed simply in terms of the radii and refractive indices alone. This is so because, in Conrady's (1929) notation, the angle of incidence I_+ at any surface depends only on I', the angle of refraction or reflection at the preceding surface, and is independent of U', the angle the ray makes with the optical axis. For monocentric systems

$$\sin I_+=(r/r_+)\sin I', \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

and the usual formula for reflexion or refraction gives

$$\sin I'=(N/N')\sin I, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

where for reflecting surfaces

$$N/N'=-1. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

The spherical aberration may be found by tracing a marginal and a paraxial ray from the same object point at a distance $L_1=l_1$ from the first surface and at the same nominal aperture $\sin U_1=u_1$. At each surface, $i=\sin I$. It readily follows that, at the last surface

$$u'_k=\sin U_1+\sin I_1\left(1-\frac{N_1}{N'_1}+\frac{N_1r_1}{N'_2r_2}+\ldots-\frac{N_1r_1}{N'_kr_k}\right), \quad \ldots\ldots \quad (4)$$

$$\begin{aligned}U'_k=U_1+I_1&-\operatorname{arsin}\frac{N_1}{N'_1}\sin I_1+\operatorname{arsin}\frac{N_1r_1}{N_2r_2}\sin I_1\\&-\operatorname{arsin}\frac{N_1r_1}{N'_2r_2}\sin I_1+\ldots-\operatorname{arsin}\frac{N_1r_1}{N'_kr_k}\sin I_1, \quad .. \quad (5)\end{aligned}$$

and the longitudinal spherical aberration is given by

$$LA'=r_k\sin I'_k(1/u'_k-1/\sin U'_k). \quad \ldots\ldots\ldots\ldots \quad (6)$$

For the later discussion it will be found more convenient to use the transverse spherical aberration

$$TA'=LA'\sin U'_k,$$

which is given by

$$TA'=r_k\sin I'_k(\sin U'_k/u'_k-1). \quad \ldots\ldots\ldots\ldots \quad (7)$$

III. CONCENTRIC SCHWARZSCHILD SYSTEM

The primary spherical aberration of a combination of a convex and a concave mirror is zero if the surfaces are concentric and of radii $\sqrt{5}+1$ and $\sqrt{5}-1$ times the focal length, as shown by Burch (1947). This result applies both to a Schwarzschild system as shown in Figure 1 and to a reversed Cassegrain system as in Figure 2. The latter system is not practicable without an auxiliary optical system as it has a virtual object distance when corrected for spherical aberration. It has been used, however, in combination with refracting elements by Bouwers (1946) and by Grey and Lee (1949).

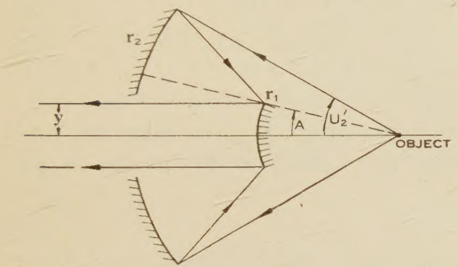


Fig. 1
Schwarzschild objective.

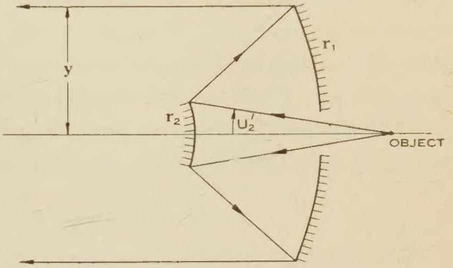


Fig. 2
Reversed Cassegrain objective.

(a) *Approximate Theory*

To apply Conrady's (1905) criteria for the best state of correction for spherical aberration we require the order of magnitude of the predominant uncorrected aberration. For this we shall consider the longer conjugate distance (the image distance for a microscope objective) to be infinite. Tracing rays from this image towards the object,

$$i_1 = \sin I_1 = y/r_1 = c_1 y,$$

where y is the aperture at the first surface and $c_1 = 1/r_1$.

Equations (4) and (5) give

$$u'_2 = 2\ (\sin I_1 + \sin I_2) = 2y(c_1 - c_2),$$

$$U'_2 = 2\ (I_1 + I_2) = 2\ (\operatorname{arsin} c_1 y - \operatorname{arsin} c_2 y),$$

and equation (7) gives

$$TA' = \frac{\sin U'_2}{2(c_1 - c_2)} - y.$$

Expanding as a power series in y, we obtain

$$\begin{aligned} TA' = & -\tfrac{1}{2} y^3 (c_1^2 - 3c_1c_2 + c_2^2) \\ & - \frac{y^5}{8(c_1 - c_2)} (c_1^5 - 8c_1^3c_2^2 + 8c_1^2c_2^3 - c_2^5) \\ & - \frac{y^7}{16(c_1 - c_2)} (c_1^7 - 4c_1^5c_2^2 + 4c_1^2c_2^5 - c_2^7) \\ & + \dots\ . \end{aligned}$$

The primary aberration in y^3 is zero when

$$c_1/c_2=r_2/r_1=(3\pm\sqrt{5})/2.$$

The focal length $f'=1/2(c_1-c_2)$, so

$$r_1=(\sqrt{5}-1)f',$$

$$r_2=(\sqrt{5}+1)f'.$$

as shown by Burch (1947).

For these radii and for unit focal length

$$TA'=-0\cdot0546y^5-0\cdot0213y^7+\ldots\ .$$

The aberration in y^5 is not corrected and for y of the order of $0\cdot6$ it is the dominant aberration. Conrady (1905) has shown that under such conditions the best correction for spherical aberration is obtained when $TA'=0$ for the marginal ray. The zonal aberration must then satisfy the condition that ΣW, the optical path difference for the marginal ray, shall be less than 2λ, and preferably less than λ as the obscuration adds to the deterioration of the image.

(*b*) *Exact Theory*

The condition $TA'=0$ for the marginal ray gives r_1 and r_2 directly for a given numerical aperture and magnification. The latter is negative for an inverted image and is denoted by $-m$, where m is positive for an inverted image and negative for an upright image.

Then

$$\sin U'=(-1/m)\sin U'_2. \quad\ldots\ldots\ldots\ldots\ldots\ldots (9)$$

Equations (4) and (5) now give

$$(U'_2-U_1)/2=I_1+I_2,$$

$$u'_2=2(\sin I_1+\sin I_2)-(1/m)\sin U'_2\,;$$

hence

$$u'_2=2\sin I_1+2\sin[(U'_2-U_1)/2-I_1]-(1/m)\sin U'_2.$$

Equation (7) gives the condition for zero spherical aberration as

$$u'_2=\sin U'_2,$$

whence

$$\cos\left(I_1-\frac{U'_2-U_1}{4}\right)=\frac{(1+1/m)\sin U'_2}{4\sin[(U'_2-U_1)/4]}. \quad\ldots\ldots (10)$$

If this condition is satisfied, the details of the optical system are found from

$$I_2=(U'_2-U_1)/2-I_1, \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (11)$$

$$r_1=2f'(\sin I_1+\sin I_2)/\sin I_1, \quad\ldots\ldots\ldots\ldots\ldots (12)$$

$$r_2=-2f'(\sin I_1+\sin I_2)/\sin I_2, \quad\ldots\ldots\ldots\ldots (13)$$

since

$$r_1/r_2=-\sin I_2/\sin I_1$$

and

$$f'=1/2(c_1-c_2).$$

The diameters of the mirrors are $2r_1\sin(U_1+I_1)$ for the convex mirror, and $2r_2\sin(U'_2-I_2)$ for the concave mirror. For an object of radius h_1 and an aperture at infinity in the object space, the diameter of the concave mirror should be increased by $2h$. There is some vignetting by a convex mirror of this size for finite objects as it is not exactly at the aperture of the objective, but it is not advisable to increase its diameter as a greater obscuration ratio would result.

(i) *Maximum Aperture.*—The maximum value of sin U'_2 for this objective is determined by the condition $\Sigma W < \lambda$. From Conrady's (1929) expression for W we obtain, for a system corrected so that $TA'=0$ for the marginal ray,

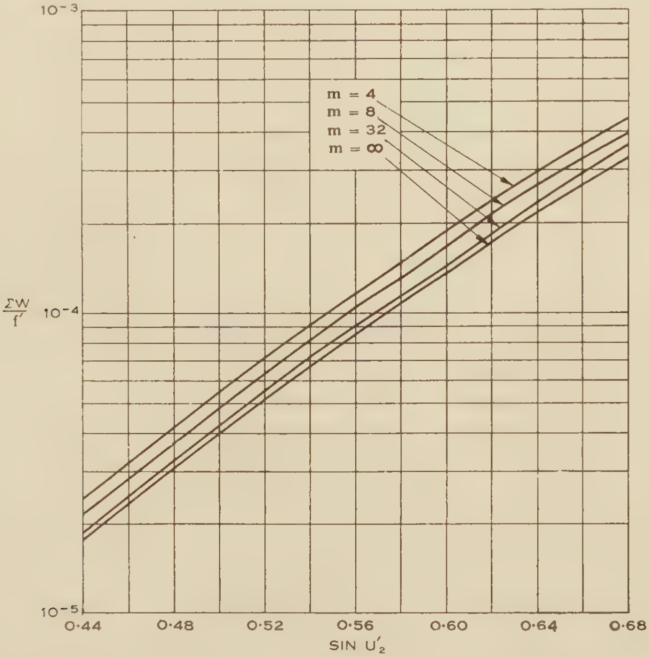


Fig. 3.—Optical path differences for corrected objectives.

$$\Sigma W = 4f' (\sin I_1 + \sin I_2)\left[\frac{\sin I_1 \sin^2 (U_1+I_1)}{(1+\cos I_1)\{1+\cos (U_1+I_1)\}\{\cos I_1+\cos (U_1+I_1)\}} + \frac{\sin I_2 \sin (U'_2-I_2)}{(1+\cos I_2)\{1+\cos (U'_2-I_2)\}\{\cos I_2+\cos (U'_2-I_2)\}}\right] \ldots \quad (14)$$

From the values of U_1, I_1, and I_2 in terms of m and U'_2 given by equations (**9**), (**10**), and (**11**), $\Sigma W/f'$ has been calculated for $m=\infty$, 32, 8, and 4, and the results are shown in Figure 3. $\Sigma W/f'$ increases slightly as m decreases in the above range, but for $m<3$ it decreases again, becoming zero at $m=1$. The latter corresponds to the well-known case of object and image at the centre of curvature of a concave spherical mirror and equation (**13**) then gives $r_2=\infty$.

The maximum permissible value of sin U'_2 is found from Figure 3, and the design of the system may then be completed using equations (**10**) to (**13**). This maximum permissible value of sin U'_2 may also be obtained to sufficient accuracy from the empirical relation

$$\sin U'_2 \doteqdot 1\cdot 317-0\cdot 22/m+0\cdot 091/(m-1)+0\cdot 186 \log_{10} \lambda/f', \quad \ldots . \; (\mathbf{15})$$

for an objective of magnification $-m$, focal length f', corrected for use at a shortest wavelength λ. This relation is based on a tolerance of λ for ΣW. If the tolerance is increased to 2λ, then sin U'_2 is increased by $0\cdot 056$.

(ii) *Under or Over Correction.*—For an objective intended for use with a cover glass or an eyepiece, the marginal spherical aberration of the objective should not be zero, but should compensate that of the rest of the system. In this case it may be shown that

$$\cos\left(I_1-\frac{U'_2-U_1}{4}\right)=\frac{\sin U'_2}{8\sin[(U'_2-U_1)/4]}\left[1+\frac{1}{m}-\frac{LA'}{f'}+\sqrt{\left(1+\frac{1}{m}-\frac{LA'}{f'}\right)^2+\frac{4LA'}{mf'}}\right]$$

$$\doteqdot \frac{(1+1/m)\sin U'_2}{4\sin[(U'_2-U_1)/4]}-\frac{TA'}{4f'(1-1/m)\sin[(U'_2-U_1)/4]}, \quad \ldots (\mathbf{16})$$

where LA' or TA' is the required aberration. To design such an objective, I_1 is found from equation (**16**) instead of equation (**10**) and the details of the system are found as before from equations (**11**), (**12**), and (**13**).

The spherical aberration of a cover glass of thickness d and refractive index N requires an under correction of the objective by an amount

$$TA'=\frac{d}{n}\cdot\frac{N\sin 2U'_3-\sin 2U'_2}{2\cos U'_3},$$

where $\sin U'_3=(1/N)\sin U'_2.$

When so corrected the residual coma of the system is given by

$$OSC'\doteqdot(1+1/m)TA'\sin U'_2/f',$$

and the longitudinal chromatic aberration expressed as a path difference (Conrady 1904), namely, $(d'-D')\delta N$, is equal to $-d(\sec U'_3-1)\delta N$.

(iii) *Immersion Objective.*—The numerical aperture of the objective may be increased by adding a front lens whose centre of curvature is at the object, and using a homogeneous immersion liquid (Burch 1947). Axial rays pass radially through the surface of this lens and the corrections of spherical aberration and of longitudinal chromatic aberration are unaltered. Both the numerical aperture and the magnification are increased by a factor equal to the refractive index of the lens. The magnification thus depends on wavelength and so the objective has some transverse chromatic aberration which will limit the range of wavelengths for which it may be used at any one time, unless this aberration is compensated in an eyepiece. The wavelength range is also limited if the dispersion of the immersion liquid differs from that of the front lens. In addition to transverse chromatic aberration the objective, being no longer monocentric, has small amounts of coma and astigmatism, but these have been found to be negligible.

When designing an immersion objective using equation (**15**), it should be noted that the wavelength to be used is that within the medium of the front lens, that is, λ_{air}/N, while f' is the focal length of the objective without the front lens.

(iv) *Obscuration.*—The convex mirror obscures the central part of the aperture of the concave mirror and the obscuration ratio is sin A/sin U'_2, where $2A$ is the angle of obscuration measured from the object. By approximate theory for $m=\infty$ (Burch 1947) this ratio is $1/\sqrt{5}$ or $0\cdot446$.

The effect of this amount of obscuration on the light distribution in the diffraction image of a point object is shown in Figure 4 for an objective with zero spherical aberration, together with that for an unobscured aperture, the scale being adjusted so that the light fluxes are equal. The decrease in diameter of the first dark ring would give slightly higher resolution on bright objects. However, the great increase in brightness of the first bright ring causes a considerable deterioration in the contrast obtained, particularly for dark objects on a bright background.

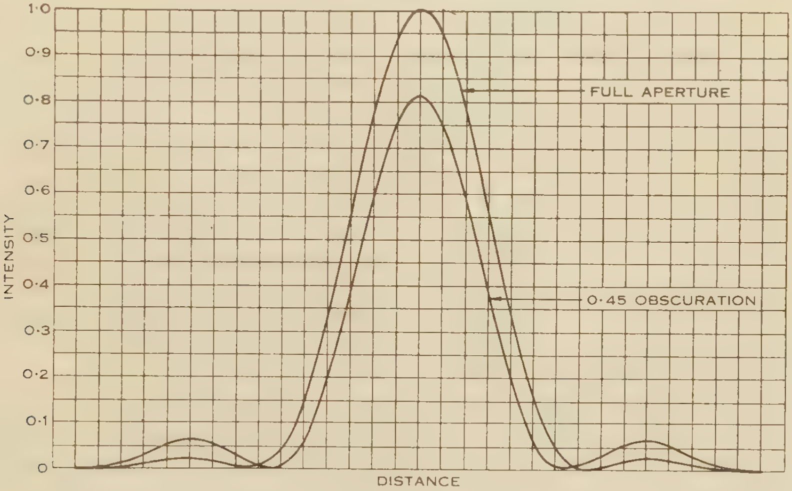


Fig. 4

IV. Objectives of Lower Obscuration

The only serious optical defect of the objective discussed in Section III is its poor contrast caused by the high obscuration. Bouwers (1946) has found that for good contrast the obscuration should not exceed $0\cdot3$. We shall therefore consider methods of reducing the obscuration to about this value.

(a) *Non-Concentric Objective*

The obscuration can be reduced by using non-concentric mirrors in a Schwarzschild system, but at the expense of introducing coma and astigmatism.

Here the obscuration $1/x$ for a mirror separation d (which in Conrady's notation is equal to $-d'_1$) is given by

$$x=(l'_2-d)/f',$$

$$=\frac{r_2(r_1+2d)-2d(r_1+2d-r_2)}{2(r_1+2d-r_2)f'}.$$

The primary spherical aberration of this system is found to be

$$LA'_p=\frac{y^2}{4r_1^2r_2(r_1+2d-r_2)^2}[(r_1+2d)^2(r_1+2d-2r_2)^2-r_1r_2^3].$$

When $LA'_p=0$,

$$r_1/2f'=\pm\sqrt{1-1/a},$$

where $a^2=r_2/r_1$, giving

$$x=a\pm\sqrt{1-1/a}. \qquad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{17})$$

For a prescribed obscuration, equation (**17**) is solved as a cubic in a. Then

$$r_1=2f'\sqrt{1-1/a},$$
$$r_2=a^2r_1,$$
$$d=a^2f'(1-1/a)+(r_2-r_1)/2.$$

The three roots of (**17**) are real and a is chosen from a consideration of the other aberrations. The value of a and the other aberrations are given in Table 1 for an object size $h=0\cdot012f'$, and for an obscuration of 1/3.

TABLE 1

a	OSC'	Astigmatism	Curvature
3·861	0·0219	0·000 018f'	0·000 078f'
2·254	0·0127	0·000 012f'	0·000 078f'
—0·115		(d negative)	

The final correction for the magnification desired must be made using ray traces, which show that the zonal spherical aberration is tolerable up to the same numerical aperture as in a concentric objective. However, serious amounts of coma are present, the OSC' being about six times the usual tolerance for objective systems (0·0025).

(*b*) *Concentric Objective*

As the non-concentric objective has serious coma, the design of a monocentric objective with lower obscuration will now be discussed. The obscuration ratio for a concentric Schwarzschild objective decreases with m, and Grey (1949) has given an approximate relationship between obscuration and magnification. The values given by his expression are shown in Table 2 and compared with $\sin A/\sin U'_2$ calculated from the relation

$$r_1 \sin (U_1+I_1-A)=(L'_2-r_2) \sin A.$$

TABLE 2

Magnification	m	∞	100	32	8	4	2	1
Obscuration (i)	Approx.	0·45	0·44	0·43	0·38	0·31	0·18	0
(ii)	Exact	0·46	0·46	0·44	0·39	0·32	0·19	0

It is seen that a magnification of magnitude 4 would give the required obscuration of about 1/3. Such a magnification is too low for most practical purposes but the objective could be used with a further concave mirror to provide useful magnification.

Such a three-mirror system can be designed with all the surfaces concentric using an extension of the theory of Section III. The final aperture $\sin U'_3$, determined by the condition $\Sigma W < \lambda$, has been found to be approximately the same as that given by equation (**15**) for two-mirror objectives. We then have

$$\sin U_1 = (1/m) \sin U'_3,$$

$$\sin U_2 = -(1/m) \sin U'_3, \qquad \textbf{(18)}$$

where $-m$ is the total magnification desired and $-m'$ is the magnification required for the first two mirrors to give the desired obscuration.

The angles of incidence are then found from

$$I_1 = (U_2 - U_1)/2, \qquad \textbf{(19)}$$

$$\cos\left(I_2 - \frac{U'_3 - U_2}{4}\right) = \left(1 - \frac{1}{m} \sin U'_3 - 2 \sin I_1\right)/4 \sin \frac{U'_3 - U_2}{4}, \qquad \textbf{(20)}$$

$$I_3 = (U'_3 - U_2)/2 - I_2, \qquad \textbf{(21)}$$

giving the radii of the three surfaces

$$r_1 = 2f' (\sin I_1 + \sin I_2 + \sin I_3)/\sin I_1, \qquad \textbf{(22)}$$

$$r_2 = -2f' (\sin I_1 + \sin I_2 + \sin I_3)/\sin I_2, \qquad \textbf{(23)}$$

$$r_3 = 2f' (\sin I_1 + \sin I_2 + \sin I_3)/\sin I_3. \qquad \textbf{(24)}$$

This objective is shown diagrammatically in Figure 5.

V. Practical Results

Several objectives of the two-mirror concentric type have been constructed for magnifications of −32 and −750 and focal length 5 mm. The numerical apertures were 0·59 dry, 0·90 oil immersion for $\lambda = 5550$ Å, and 0·78 for $\lambda = 2750$ Å using a fused quartz front lens and a homogeneous immersion liquid of the type described by Martin and Johnson (1928).

The spherical aberration of each of these objectives, when adjusted, was found to be well corrected, and the diffraction image of a point object agreed closely in appearance with the result shown in Figure 4. The performance on high contrast objects is excellent, but with low contrast objects or objects with a bright background the image is noticeably inferior to that given by a refracting objective.

The two mirrors for the dry objectives are separately mounted and have independent adjustment for centring and for separation. The convex mirror for the immersion objectives is cemented to the front lens, considerably simplifying the construction. No difficulty has been experienced in centring it to the lens surfaces. The separation of the mirrors is adjusted to give least spherical aberration at the image of an artificial star object; at the same time the centring is adjusted to remove coma.

A non-concentric objective has been made for N.A. $=0\cdot78$ but has not been successfully adjusted. The combination of the coma inherent in the design with coma due to misalignment of the mirrors gives a diffraction pattern that is difficult to correct, and the writer has been unable to devote the necessary time to a systematic adjustment of this system.

The three-mirror objective shown in Figure 5 has not yet been constructed. As may be seen, an oblique plane mirror is required to bring the light out again, so the objective could not be used in a conventional microscope. The obscuration due to this plane mirror cannot be made to coincide with the image of the initial obscuration, but if the object is small the vignetting should not be serious.

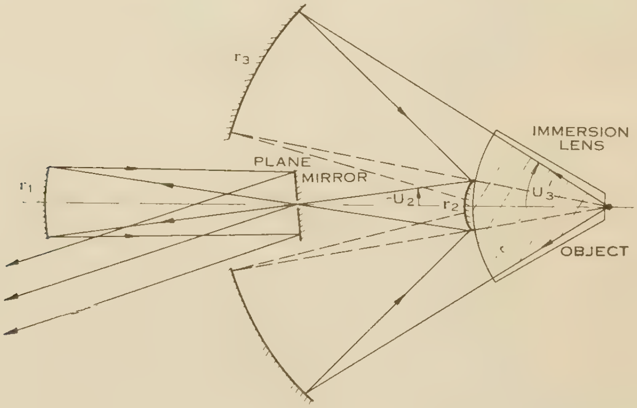


Fig. 5.—Three-mirror objectives.

VI. Conclusions

The reflecting objectives discussed have satisfied the two conditions considered necessary for simple construction and ease of use outside the visible spectrum : they employ only spherical surfaces, and their foci are independent of wavelength due to the absence of refracting surfaces at other than normal incidence.

Many workers have found that the concentric Schwarzschild objective, although yielding low contrast, is a satisfactory system. It has been shown here that such a concentric system is unique optically as it can be designed by directly solving an exact equation to satisfy the condition that the marginal spherical aberration is zero. The equations given here enable such an objective to be designed in a few minutes.

For application where the low contrast would be serious, the theoretical discussion presented here indicates that the three-mirror objective could be an improvement. The alternative of a non-concentric objective is not as promising since the large amount of coma restricts its use to very small objects and makes it difficult to adjust.

VII. ACKNOWLEDGMENTS

The author wishes to acknowledge the assistance of the other members of the Division of Physics, C.S.I.R.O.; Mr. V. R. Schaefer who constructed these objectives, Miss J. Ward who performed the necessary computations, and Miss E. M. Philpott who assisted with their adjustment and the preparation of immersion liquid.

VIII. REFERENCES

BOUWERS, A. (1946).—" Achievements in Optics." (Elsevier Publishing Co.: New York and Amsterdam.)

BRUMBERG, E. M. (1941).—*C.R. Acad. Sci. U.R.S.S.* **23**: 486.

BRUMBERG, E. M. (1943).—*Nature* **152**: 357.

BURCH, C. R. (1947).—*Proc. Phys. Soc. Lond.* **59**: 41.

CONRADY, A. E. (1904).—*Mon. Not. R. Astr. Soc.* **64**: 182.

CONRADY, A. E. (1905).—*Mon. Not. R. Astr. Soc.* **65**: 594.

CONRADY, A. E. (1929).—" Applied Optics and Optical Design." (Oxford Univ. Press.)

GREY, P. S. (1949).—*J. Opt. Soc. Amer.* **39**: 723.

GREY, P. S. (1950).—*J. Opt. Soc. Amer.* **40**: 283.

GREY, P. S., and LEE, P. H. (1949).—*J. Opt. Soc. Amer.* **39**: 719.

JOHNSON, B. K. (1949).—*J. Sci. Instrum.* **26**: 148.

MARTIN, L. C., and JOHNSON, B. K. (1928).—*J. Sci. Instrum.* **5**: 380.

SEEDS, W. E., and WILKINS, M. H. F. (1949).—*Nature* **164**: 228.

SCHWARZSCHILD, K. (1905).—" Theorie der Spiegelteleskop." (Observatory: Göttingen.)

SOLUTION OF A HEAT FLOW PROBLEM

By J. R. M. Radok*

[*Manuscript received June 22, 1950*]

Summary

This paper deals with the heat flow through a rectangle subject to the following boundary conditions: One end is completely insulated (without heat flux across it) and at the opposite end a constant temperature gradient is maintained; the remaining sides radiate into a medium which is at zero temperature. Initially the rectangle is at zero temperature. The problem is converted into a homogeneous one by considering a rectangle of twice the length with a uniformly distributed heat source along the centre line. The solution of this problem is effected by using particular solutions given in Carslaw and Jaeger.† The problem considered is of interest in connection with the testing of heat-insulating materials.

I. Introduction

The thermal properties of insulating materials may be determined experimentally by applying a variable state method to rectangular blocks. The method involves measurement of the temperature on one face of the block across which the heat flow is zero while a constant temperature gradient is maintained across the opposite face. The remaining four faces are allowed to radiate into a medium assumed at zero temperature. In actual cases it is often possible to assume two-dimensional conditions, so that one has only to deal with the problem of heat flow in a rectangle two opposite sides of which are radiating outwards.

In this paper the theoretical solution of the two-dimensional problem is given to enable experimental investigators to estimate radiation losses. The method used is equally applicable to the problem in one dimension, when no radiation takes place, or in three dimensions with four radiating sides. It makes use of special solutions of the heat equation which in the terminology of Carslaw and Jaeger† correspond to "continuous" heat sources.

II. Formulation of the Problem

In two dimensions the differential equation of conduction of heat is

$$\frac{\partial v}{\partial t}=\varkappa\left\{\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}\right\}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where v is the temperature, t is the time, x and y are rectangular coordinates,

and
$$\varkappa=\frac{K}{c\rho},$$

---

* Aeronautical Research Laboratories, Department of Supply, Melbourne.

† Carslaw, H. B., and Jaeger, J. C. (1947).—"Conduction of Heat in Solids." (Oxford Univ. Press.)

K being the thermal conductivity, c the specific heat, and ρ the density of the conducting material. For the rectangular region, given by Figure 1, the initial and boundary conditions relevant to the above problem are

$$v(x, y, t)=0 \text{ for } t=0, \qquad \text{(2)}$$

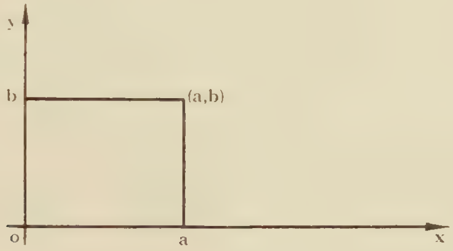


Fig. 1

that is, at the beginning of the experiment the specimen is assumed at zero temperature.

$$\partial v/\partial x=0 \text{ at } x=0, \qquad \text{(3)}$$

that is, there is no flux of heat across the insulated side.

$$\partial v/\partial y-\chi v=0 \text{ at } y=0, \qquad \text{(4)}$$

$$\partial v/\partial y+\chi v=0 \text{ at } y=b, \qquad \text{(5)}$$

that is, outwards radiation takes place proportional to the temperature of the indicated boundaries provided χ, which is sometimes referred to as exterior conductivity, is positive.

$$\partial v/\partial x=\varphi/K \text{ at } x=a, \qquad \text{(6)}$$

that is, the temperature gradient across this side is constant.

Carslaw and Jaeger (1947, p. 22) show that the solutions of two- and three-dimensional problems can be obtained as the product of linear solutions, provided the boundary conditions are homogeneous. Although the condition (6) above does not satisfy this requirement, it is possible to transform the present problem in such a way that a product solution is possible. For this purpose the boundary, shown in Figure 1, will be replaced by the one indicated in Figure 2, while only the boundary condition (6) will be changed as follows:

$$\partial v/\partial x=0, \text{ at } x=2a, \qquad \text{(7)}$$

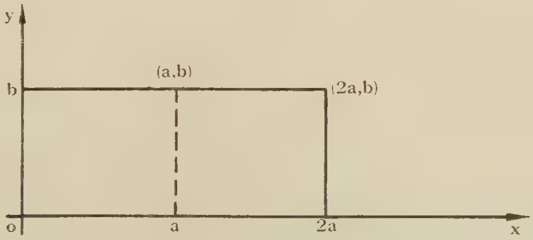


Fig. 2

and continuous sources of constant strength $2\varphi/\rho c$ distributed uniformly on the side (a, o) to (a, b) (**8**). In this way the centre line of the rectangle of Figure 2 becomes a line of thermal symmetry with equal amounts of heat flowing in the positive and negative x direction.

III. SOLUTION OF THE PROBLEM

The solution of this problem will be obtained as the product of the solutions of the following two linear problems with the initial condition (**2**) and the boundary conditions :

(i) (**3**) and (**7**) and an instantaneous source $2\varphi/\rho c$ at $x=a$,

(ii) (**4**) and (**5**) and an instantaneous unit source at $y=y'$.

From page 298 of Carslaw and Jaeger the solution of (i) is

$$v_e(x, t)=\frac{1}{a}\{\tfrac{1}{2}+\sum_{n=1}^{\infty} \cos \tfrac{1}{2}\, \nu_n x \cos \tfrac{1}{2}\pi n e^{-(\frac{1}{2}\nu_n)^2 \varkappa t}\},$$

$$=\frac{1}{a}\{\tfrac{1}{2}+\sum_{n=1}^{\infty} (-1)^n \cos \nu_n x e^{-(\nu_n)^2 \varkappa t}\}, \quad \dots\dots\dots\dots \quad (\mathbf{9})$$

where

$$\nu_n=\frac{n\pi}{a},$$

and the solution of (ii) is

$$v_e(y, t)=2 \sum_{m=1}^{\infty} \frac{\{\mu_m \cos \mu_m y+\chi \sin \mu_m y\}\{\mu_m \cos \mu_m y'+\chi \sin \mu_m y'\}}{b\{\mu^2{}_m+\chi^2\}+2\chi} e^{-\varkappa(\mu_m)^2 t},$$

$$\dots\dots\dots\dots\dots\dots \quad (\mathbf{10})$$

where

$$\tan b\mu_m=\frac{2\mu_m\chi}{\{\mu^2{}_m-\chi^2\}}. \quad \dots\dots\dots\dots\dots\dots \quad (\mathbf{11})$$

Hence it follows that the solution of the two-dimensional problem is given by

$$v=\frac{4\varphi}{\rho ca}\int_0^t \int_0^b \{[\tfrac{1}{2}+\sum_{n=1}^{\infty} (-1)^n \cos \nu_{n\,e} x e^{-\nu_n{}^2 \varkappa(t-\tau)}]$$

$$\times \left[\sum_{m=1}^{\infty} \frac{(\mu_m \cos \mu_m y+\chi \sin \mu_m y)(\mu_m \cos \mu_m y'+\chi \sin \mu_m y')e^{-\varkappa\mu_m{}^2(t-\tau)}}{b(\mu^2{}_m+\chi^2)+2\chi}\right]\Big\} dy' d\tau$$

$$=\frac{4\varphi}{\rho ca}\Big\{\tfrac{1}{2}\sum_{m=1}^{\infty} (\mu_m \cos \mu_m y+\chi \sin \mu_m y)\left[\sin \mu_m b-\frac{\chi}{\mu_m}(\cos \mu_m b-1)\right] \frac{[1-e^{-\mu^2{}_m \varkappa t}]}{\varkappa\mu^2{}_m[b(\mu^2{}_m+\chi^2)+2\chi]}$$

$$+\sum_{n=1}^{\infty} (-1)^n \cos \nu_n x \sum_{m=1}^{\infty} [\mu_m \cos \mu_m y+\chi \sin \mu_m y]\left[\sin \mu_m b-\frac{\chi}{\mu_m}(\cos \mu_m b-1)\right] \frac{1-e^{-(\mu^2{}_m+\nu^2{}_n)\varkappa t}}{\varkappa(\mu^2{}_m+\nu^2{}_n)[b(\mu^2{}_m+\chi^2)+2\chi]}\Big\}.$$

$$\dots\dots\dots\dots\dots\dots\dots \quad (\mathbf{12})$$

But by (**11**)

$$\sin \mu_m b=\pm\frac{2\mu_m\chi}{(\mu^2{}_m+\chi^2)}, \text{ for } m \begin{matrix}\text{odd}\\ \text{even}\end{matrix} \quad \dots\dots\dots\dots \quad (\mathbf{13})$$

and

$$\cos \mu_m b = \pm \frac{(\mu^2{}_m - \chi^2)}{(\mu^2{}_m + \chi^2)}, \text{ for } m \begin{matrix} \text{odd} \\ \text{even} \end{matrix} \quad \ldots\ldots\ldots\ldots \quad (14)$$

whence

$$\sin \mu_m b - \frac{\chi}{\mu_m} (\cos \mu_m b - 1) = \begin{matrix} 2\chi/\mu_m \\ 0 \end{matrix}, \text{ for } m \begin{matrix} \text{odd} \\ \text{even} \end{matrix} \quad \ldots\ldots \quad (15)$$

Using (**15**) the solution (**12**) may be written

$$v = \frac{8\varphi\chi}{aK} \sum_{m\,\text{odd}}^{\infty} \frac{\cos \mu_m y + (\chi/\mu_m) \sin \mu_m y}{b(\mu^2{}_m + \chi^2) + 2\chi} \left\{ \frac{1 - e^{-\mu^2{}_m \varkappa t}}{2\mu^2{}_m} + \sum_{n=1}^{\infty} (-1)^n \cos \nu_n x \, \frac{1 - e^{-(\mu^2{}_m + \nu^2{}_n)\varkappa t}}{\mu^2{}_m + \nu^2{}_n} \right\}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (16)$$

Using (**12**) or (**15**), the temperature v at any point inside the rectangular region under consideration may be found at any time t after the start of the experiment.

THE MOMENTUM SPECTRUM AND POSITIVE EXCESS OF COSMIC RAY MESONS AT SEA-LEVEL

D. E. Caro,*† J. K. Parry,* and H. D. Rathgeber*

[*Manuscript received September 29, 1950*]

Summary

A spectrometer for momentum measurements on cosmic ray particles is described. Charged C.R. particles are deflected in the air gap of an electromagnet and the deflection is recorded by an array of G.M. counters. The momentum distribution and positive excess of particles which penetrate 10 cm. of lead is determined over a momentum range $2 \cdot 4 \times 10^8$ eV./c. to 5×10^{10} eV./c. from measurements on 6056 particles.

I. Introduction

Previous attempts to measure the deflection of cosmic ray particles by a magnetic field in air using G.M. counter techniques have met with only small success (Curtiss 1932 ; Rathgeber 1938) due mainly to insufficient resolving power. A reconsideration of the design requirements in the light of the results obtained with cloud chambers (Blackett 1937 ; Jones 1939 ; Hughes 1940 ; Wilson 1946) led to the conclusion that the counter method was worth further investigation.

As a result it was found to be practicable to build equipment for measuring the momentum of cosmic ray particles which would exceed the resolving power of previous cloud chamber apparatus. It would also be capable of collecting information at a much higher rate although the detailed information given by the cloud chamber method would be lost. However, by using a large number of G.M. counters it would be possible to identify the track of a single particle and also to detect practically all showers (Rathgeber 1938).

II. Description of Equipment

(*a*) *General Description*

When G.M. counters are used to indicate the path of a particle, the instrument can be considered as the analogue of an optical system consisting only of slits. The counters in trays 1 and 2 (Fig. 1) serve to define the track of the particle prior to its entry into the magnetic field. The deflection received by the particle in traversing the field is determined by the track position in tray 3.

The resolving power is limited by the diameter of the counters, the lower limit being set by the difficulty of constructing reliable counters with a diameter smaller than 1 cm. This limitation has been overcome by the use of overlapping counters. The passage of a particle through the area common to adjacent

---

* Department of Physics, University of Melbourne.

† Now at University of Birmingham.

counters is indicated by their simultaneous discharge and is thus detectable as a separate event from the discharge of one counter alone. In the present case the use of two layers of counters reduces the effective slit width by a factor three.

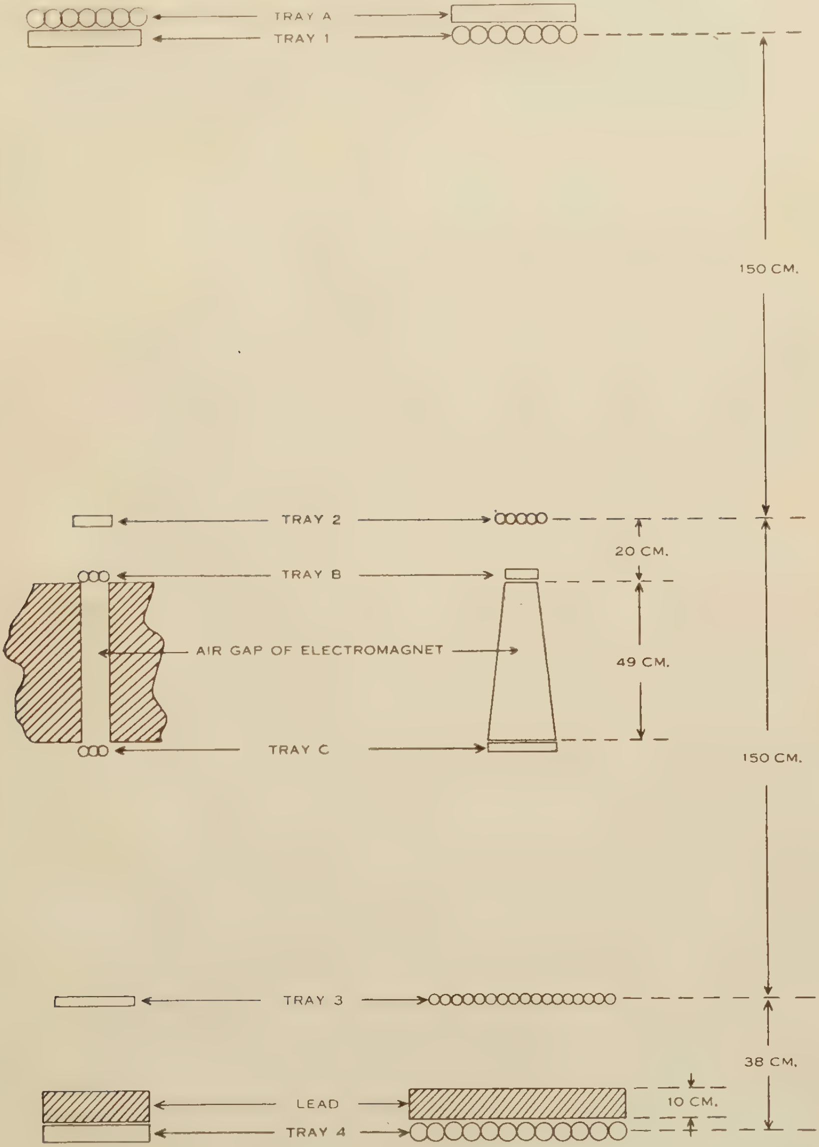

TRAY A
TRAY 1
150 CM.
TRAY 2
20 CM.
TRAY B
AIR GAP OF ELECTROMAGNET
49 CM.
TRAY C
150 CM.
TRAY 3
38 CM.
LEAD
10 CM.
TRAY 4

Fig. 1.—Layout of magnet and counter trays. (Counter trays not to scale. For details see Table 1.)

The momentum of a singly charged particle which receives a deflection D as measured in tray 3 is given by the relation

$$p=(1\cdot 75\times 10^{6})\, B/D,$$

where B=magnetic flux density in gauss,

p=momentum in units eV./c.,

D=deflection in cm.

The numerical constant has been evaluated by taking into account the geometry of the system and the effect of the fringing field. This is discussed in more detail in a later section of the paper. Operating at the maximum field of 13,500 gauss and an effective slit width in tray 3 of 0·45 cm. the maximum measurable momentum is 5×10^{10} eV./c.

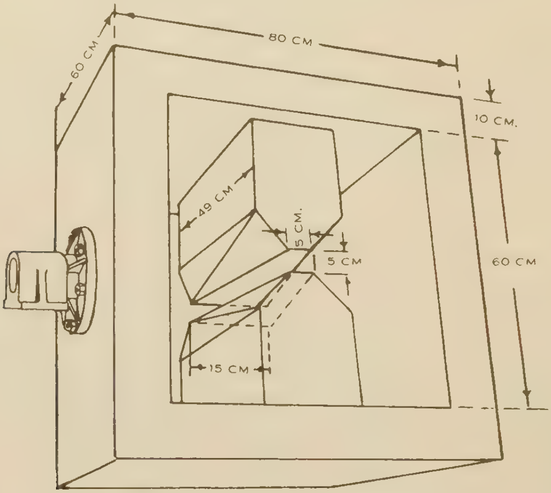


Fig. 2.—Iron core of the electromagnet.

To prevent the recording of particles other than those taking suitable paths through the magnetic field, additional trays A, B, and C are used. Trays A, B, and C are connected in a threefold " master " coincidence circuit arranged so that the counters in trays 1, 2, and 3 record only when the particle has set off a threefold coincidence. This method of " master pulses " has previously been used in conjunction with photographic recording (Swann 1940 ; Regener 1943). It was realized that the analysis of a large number of such photographs (over 200 per day at low magnetic fields) would be a very lengthy process. In consequence a punched card system of recording was used. Telephone selector switches and an automatic form of punch card machine make it possible to record all the information obtained from the counters as well as the time, date, etc. directly on the card. The record cards (one per particle) are sorted after each run by an automatic sorting machine.

Tray 4 (Fig. 1) records particles which penetrate 10 cm. of lead thus separating the soft and penetrating components.

The magnet and counter assembly is mounted in a rack which is free to pivot about a north-south axis. Plate 1 shows a general view of the complete equipment. The whole apparatus is housed in a hut with very little absorbing material (approximately 1·5 g. cm.$^{-2}$) in the roof and walls.

(*b*) *The Electromagnet*

The iron yoke of the 3-ton electromagnet is shown, together with all relevant dimensions, in Figure 2. By cutting away the corners of the pole pieces, the maximum field is confined to the cross section defined by the counter trays. The total flux is thus reduced and the cross section of the iron yoke required to prevent saturation is kept small. The magnet coils take the form of 12 pancake windings fitted over each pole piece.

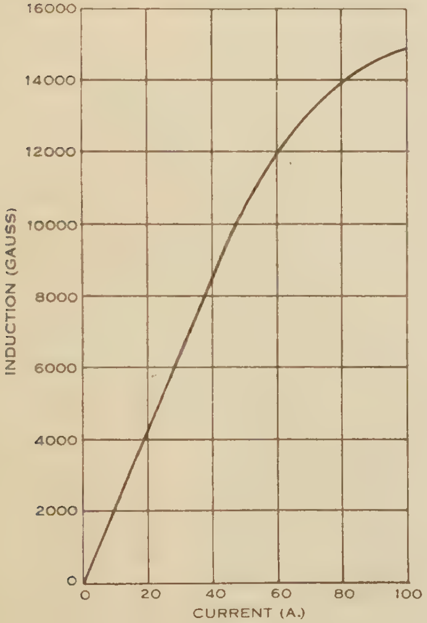


Fig. 3

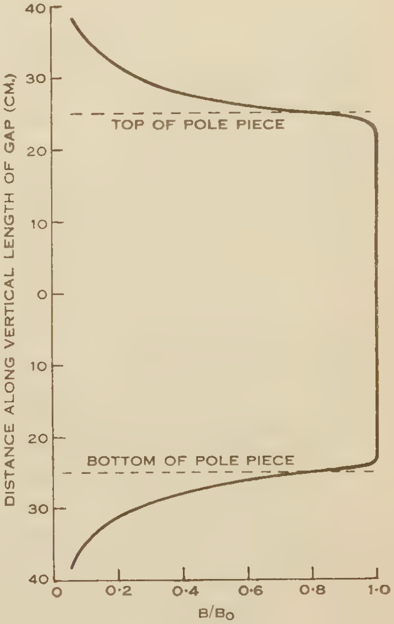


Fig. 4

Fig. 3.—Excitation curve of the electromagnet.

Fig. 4.—Distribution of magnetic induction along length of pole piece.

Figure 3 gives the magnetization curve. The points used in obtaining this curve were calculated from the e.m.f. developed between the centre and periphery of a small silver disk spun at constant speed about an axis parallel to the field (Cork and Wadey 1942). The magnet and windings were calculated to operate continuously at 10,000 gauss with natural convection and radiation cooling only. At the excitation required for this field the magnet ran with a temperature rise of only 20 °C. However, by using a forced air-cooling system, the maximum exciter current possible for continuous operation was increased to 80 A. It may be seen from the magnetization curve that only a small gain in field would be realized by operating at a current exceeding 75 A. The magnet was, therefore, operated at an induction of 13,500 gauss. The field distribution curve (Fig. 4) was obtained with a flip coil of 1 cm. diameter. As may be seen from the curve, the field is constant to within the accuracy of measurement over most of the length of the pole pieces. It begins to fall off approximately 1 cm. inside the edge of the pole pieces, falling to 10 per cent. at a distance of 10 cm. from the poles.

The current for the electromagnet is supplied by a motor generator set. This is stabilized to within 0·05 A. over the working range from 5 to 100 A. This stabilizer has been described in detail elsewhere (Caro and Parry 1949).

(*c*) *The G.M. Counters*

The sizes of the counters used in the various trays are given in Table 1 and the geometrical arrangement of the trays in Figure 1. The considerations leading to the choice of the diameters of the analysing counters have been reported previously (Rathgeber 1938).

TABLE 1

DETAILS OF COUNTER TRAYS

Tray	Number of Counters	Diameter (cm.)	Effective Length (cm.)	Size of Counting Unit (cm.)	Number of Counting Units
1	15	2·7	30	0·9	27
2	7	1·35	6	0·45	11
3	43	1·35	18	0·45	83
4	22	3·2	24	—	—
A	13	2·7	26	—	—
B	5	1·35	5	—	—
C	5	1·35	18	—	—

The envelopes of the counters were made of Pyrex glass, the cathode being deposited by evaporation of gold from the central wire.

In the present application this design has several advantages. The position of the sensitive area of the counter is well defined and it is therefore a simple matter to place them in their correct position. The material of the walls, between 0·5 and 1·0 mm. in thickness, is light and of low atomic number so that the scattering and absorption of the cosmic ray particles is small.

The counters were at first filled with an argon-ethyl alcohol mixture in the approximate proportion 10 : 1, and later with an argon-ethylene mixture in the same proportions. When filled, the pressure in the counters was adjusted to such a value that the starting voltage (pulses of 1 to 3 volts) was 1000 ± 10 V.

The counters are all operated from the same electronically stabilized, R.F. high voltage supply adjusted to 1,100 V.

(*d*) *Circuit Description*

The operation of the recording circuits may be represented by the block diagram of Figure 5. The negative counter pulses from trays *A*, *B*, and *C* are fed to a threefold Rossi coincidence circuit. The output pulse from the Rossi circuit triggers the "master pulse" generator which produces a rectangular, positive pulse of 10 μsec. duration at an impedance level of 30 ohms whenever a particle traverses trays *A*, *B*, and *C*. The resolving time of the master circuits is of the order of 3 or 4 μsec.

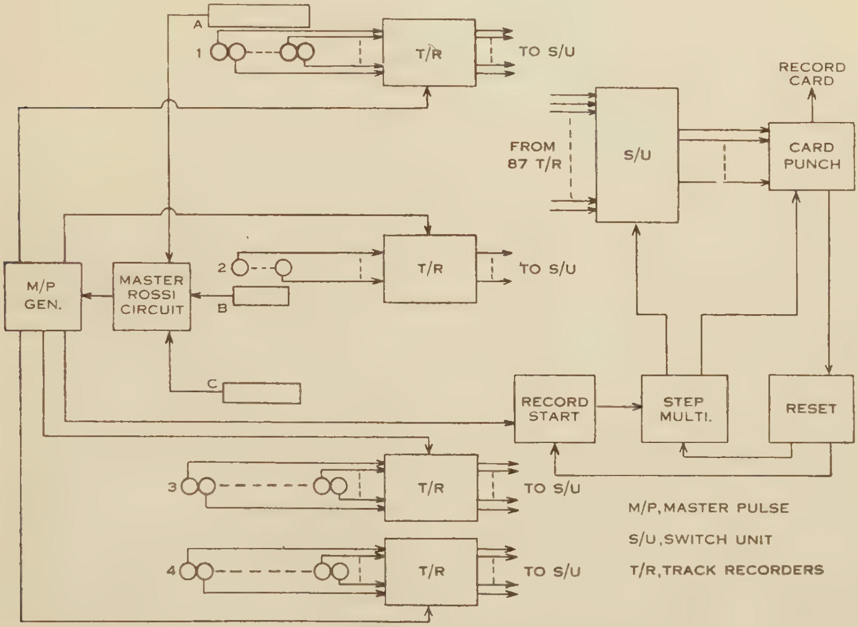

A
1
T/R
TO S/U
RECORD CARD
FROM 87 T/R
S/U
CARD PUNCH
2
T/R
TO S/U
M/P GEN.
MASTER ROSSI CIRCUIT
B
C
RECORD START
STEP MULTI.
RESET
3
T/R
TO S/U
M/P, MASTER PULSE
S/U, SWITCH UNIT
4
T/R
TO S/U
T/R, TRACK RECORDERS

Fig. 5.—Block diagram of recording equipment.

To each of the counters in trays 1 to 4 is connected a shield grid thyratron in the coincidence circuit of Figure 6. The shield grid is held negative by the master pulse line and so the valve is normally non-conducting. If a particle passes through trays *A*, *B*, and *C*, a master pulse is produced. One of two conditions will then result at the circuit of Figure 6.

(i) If the counter connected to V1 fires, the valve receives a negative pulse on its control grid simultaneously with a positive master pulse on its shield grid. The valve therefore remains non-conducting.

(ii) If the counter connected to V1 does not fire, the valve becomes conducting due to the positive master pulse on its shield grid.

Thus, immediately after a particle has passed through the system, track recorder thyratrons connected to counters through which the particle has passed have a high plate voltage while all others are conducting and consequently have a low plate voltage.

This information is transferred to the record card by the remaining circuits in Figure 5. Each counter is allotted a specified position on the card corresponding to its position in the equipment. The card machine used is a 45 column

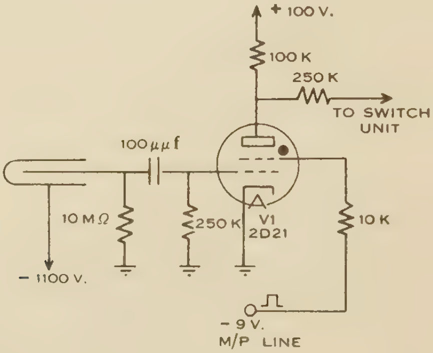

+ 100 V.
100 K
250 K
TO SWITCH UNIT
100 μμf
10 MΩ
250 K
V1
2D21
10 K
− 1100 V.
− 9 V.
M/P LINE

Fig. 6.—Track recorder circuit. One of these valves to each counter in trays 1–4.

Powers automatic card punch suitably modified to operate electrically. The contacts of a 12 bank, 25 position stepping switch (3 PMG type uniselectors, electrically ganged) are connected to the anodes of the track recorder thyratrons and to any other data to be recorded. This switch unit is wired so that its 25

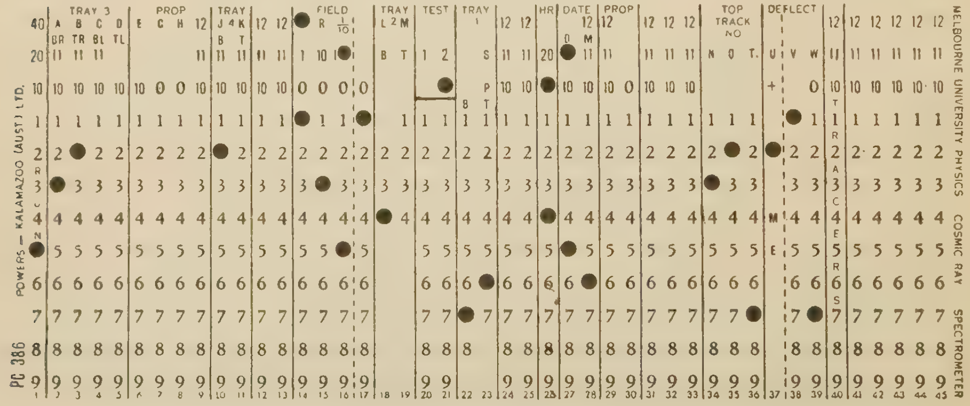

TRAY 3
PROP
TRAY
FIELD
TRAY
TEST
TRAY
HR
DATE
PROP
TOP TRACK NO
DEFLECT
COUNTERS
POWERS — KALAMAZOO (AUST.) LTD.
PC 386
MELBOURNE UNIVERSITY PHYSICS COSMIC RAY SPECTROMETER

Fig. 7.—Typical record card.

positions correspond to 25 predetermined columns on the card, each bank of the switch corresponding to one of the 12 possible positions in each column (see Fig. 7).

When a master pulse is produced, it operates the " record start " circuit. This disables the master circuits thus rendering the equipment insensitive to further tracks until the recording process is completed. The " record start " circuit also applies the H.T. to the " step multivibrator ". The function of this circuit is to step the switch unit and punch machine in unison, one position at a time with a frequency of about 1 cycle per second. Whenever a wiper of the switch moves over a contact which is at a high potential, corresponding to a counter fired in the equipment, the appropriate punch is set in the machine. When the 25th position is reached, the carriage of the punch machine closes the switch operating the " reset " circuit which resets all the recording equipment. The machine also punches and ejects the card, at the same time resetting itself.

This recording sequence takes approximately 20 seconds after which the instrument is ready to record the next track.

III. Treatment of Results

(*a*) *Analysis and Classification of Tracks*

The analysis of the results is complicated by the variety of tracks which a particle of a certain momentum may follow. As it would be impossible in a paper of reasonable length to give a detailed account of this analysis, the following section briefly outlines the method of approach.

The system of overlapping the counters in the resolving trays was mentioned in an earlier section of the paper. Figure 8 (A) shows the arrangement of the counters in one such tray, and it is seen that a tray of n counters, overlapped by $\frac{1}{3}$ of the diameter, defines $(2n-3)$ " counting units " (abbreviated to C.U. in the following text). It is convenient to represent the counter trays by the equivalent system of counting units. This is done in the simplified diagram of the track recording trays (Fig. 8 (B)).

The effective region in which the particles are deflected is not defined exactly by the boundaries of the pole pieces but it is extended by the fringing field. It is necessary to make a correction by assuming virtual boundaries which define an equivalent deflecting field, homogeneous within the region but terminated abruptly at the limits. This is accomplished readily (Coggeshall 1947) if the variation in field strength along various paths through the field is known, e.g. Figure 4. The virtual boundaries of the field found in this manner are used in the analysis.

The information obtained from the record card gives the particular C.U. in each of the three trays through which the particle passed, thus defining the track as in Figure 8 (B). It will be assumed here that the particle passes through the centre of each C.U., the spread introduced due to the finite size of the C.U. being discussed in a later section. Since the particles travel in straight lines prior to entering the field, each combination of C.U. in tray 1 and tray 2 defines a " top track ". Each top track is allotted values of α and g (see Fig. 8) which define the C.U. in tray 3 that the particle would traverse if undeflected. Since the C.U. through which the deflected particle passed is known, the actual deflection D is determined.

The following expressions have been derived relating ρ, the radius of curvature in the field, with D the deflection.

$$\left.\begin{aligned}\rho&=L(2b-L)/2D'\\ D'&=D+\delta D, \qquad \delta D=F(\alpha, g, D),\end{aligned}\right\}, \qquad \ldots\ldots\ldots\ldots \quad (1)$$

where D' is the "effective deflection" and b and L are constants of the equipment shown in Figure 8 (B).

The above relations result from geometrical considerations and involve several approximations each of which has been made to preserve a final accuracy of 1 per cent.

Remembering that

$$p=300\,B\rho, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

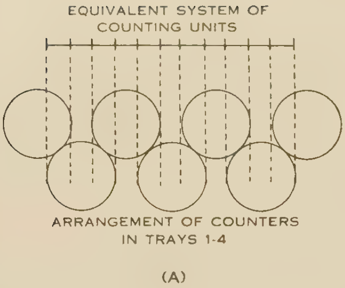

EQUIVALENT SYSTEM OF COUNTING UNITS
ARRANGEMENT OF COUNTERS IN TRAYS 1-4
(A)

Fig. 8 (A).—Arrangement of counters in trays 1–4.

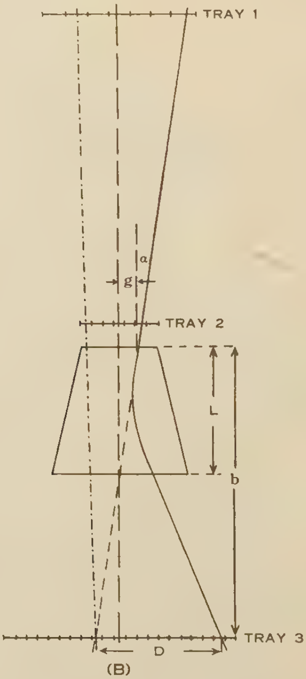


Fig. 8 (B).—Diagrammatic representation of resolving counter trays.

and expressing D' etc. in terms of the C.U. in tray 3 (one C.U. $=0\cdot45$ cm.) and B in gauss, we obtain

$$p=\frac{3\cdot9B}{D'}\ \text{MeV./c.} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

The following conclusions may be drawn from the analysis :

(i) Since D is measured to the nearest C.U., the value of δD need be known only to within one C.U. For all values of α and g encountered in the equipment, $\delta D=0$ for $D<48$ C.U. Although deflections up to 69 C.U. may be recorded for some top tracks the greatest deflection used in the results is 41 C.U. We may therefore write (3) as

$$p=\frac{3\cdot9B}{D}\ \text{MeV./c.}$$

(ii) There are 297 top tracks allowed by the counter system. Of these 96 would result in a path which, for certain deflections, may pass outside the homogeneous region of the magnetic field. These top tracks are excluded from the analysis.

(iii) The C.U. in tray 3 through which each top track would pass if undeflected may be calculated from the known α and g values. It is found that the remaining 201 top tracks pass through 29 C.U. in tray 3. Numbering the C.U. in tray 3 from 1 to 83, the undeflected top tracks traverse the alternate numbers 14, 16, 18, ... 68, 70. It may be seen from Figure 8 (B) that more than one top track may pass through the same C.U. in tray 3, and Table 2 gives the distribution of these top tracks over the alternate C.U. 14 to 70. Furthermore, the variation in α and g of the top tracks in each of these groups is so small that the effect on the radius of curvature may be neglected. Thus the 201 top tracks may be classified into 29 groups, each group being allotted a "top track number" (T) which is determined by the number of the C.U. in tray 3 which the group would traverse if undeflected.

TABLE 2

DISTRIBUTION OF TOP TRACKS

Top Track Group	Number of Top Tracks in Group	Top Track Group	Number of Top Tracks in Group	Top Track Group	Number of Top Tracks in Group
14	1	34	9	54	10
16	2	36	9	56	9
18	3	38	10	58	8
20	4	40	9	60	7
22	6	42	9	62	6
24	7	44	9	64	4
26	8	46	10	66	3
28	9	48	9	68	2
30	10	50	9	70	1
32	9	52	9		

The procedure adopted in the actual sorting of the cards now follows in a straightforward manner. A brief survey of the record cards is made and all single tracks are selected, electrons being differentiated from mesons by tray 4. The remainder, which are caused by showers and particles failing to fire one of the resolving trays, e.g. particles missing tray 3 due to magnetic deflection, are rejected for the purpose of the analysis.

The cards are then fed into a sorting machine followed by a set of "tracer cards". There are 201 tracer cards each punched to correspond to one of the allowed top tracks and distinguished from the record cards by their colour. The cards are then sorted systematically on columns 23, 22, 19, and 18 (see Fig. 7) and it is found that the tracer cards have filtered through the stack of record cards. All cards beneath a tracer card and above the next coloured card have the same top track as that of the tracer card. Similarly a set of " reject cards " is used to select the top tracks not allowed by the analysis. The 201 sets of top tracks are then reclassified into 29 groups as explained earlier, the appropriate top track number punched on each card, and the tracer and reject cards removed.

The process is repeated, this time sorting on tray 3, i.e. columns 2 to 5, and the cards are obtained in 83 groups. Each of these groups is allotted a " deflection number " (N) corresponding to the C.U. which has been discharged by the deflected particle.

The cards are now grouped according to their deflection number (N). Since the top track number (T) is already punched on the card, the deflection, $D=|T-N|$ may be obtained immediately for each card. The sign of D is determined by the direction of deflection and the direction of the magnetic field. The cards are classified into the 83 deflection groups, from $D=-41$ to $D=+41$ C.U., and the deflection punched on each card.

It is now possible to plot the distribution in deflection of the mesons incident upon the equipment. This distribution, which depends on the value of the magnetic field, may be converted to the momentum distribution by means of the relation derived previously

$$p=\frac{3 \cdot 9B}{D} \text{ MeV./c.}$$

(b) *Factors Affecting the Accuracy of the Determination*

Before presenting the results, it is necessary to consider the effect of the instrument on the measurement of the momentum distribution.

(i) *Distortion of the Deflection Spectrum due to Magnetic Cutoff.*—This first instrumental effect may be explained in the following manner. Consider the direction of the magnetic field such that positive particles are deflected towards C.U. 1 in tray 3. Of the positive particles traversing the top track group 14, those with $D>|1-14|$, i.e. >13 C.U. will be deflected away from tray 3 and consequently not recorded. For progressively higher values of the top track group this cutoff deflection increases, until for the other extreme top track group 70, the cutoff deflection is 69 C.U. A similar effect takes place for negative particles, starting with a cutoff deflection of 13 C.U. for the top track group 70. Consequently those particles with deflections up to ± 13 C.U. are recorded in their correct proportions, but for greater deflections a progressively smaller fraction of the true number of particles is recorded.

This magnetic cutoff effect has been recognized in counter controlled cloud chamber work. The result is that in the low momentum region the measured momentum distribution falls below that determined by other methods (Rossi 1948). While it is very difficult to allow for this distortion of the distribution by a cloud chamber, it is possible to calculate a set of correction factors for the equipment used in this experiment.

It has been verified experimentally that each of the 201 top tracks, apart from statistical fluctuations, contributes the same amount to the total intensity recorded by the equipment. Consider now those particles with a momentum such that they would suffer a deflection of 14 C.U. Those traversing the top track group 14 are deflected away while all other top tracks contribute to the number of particles recorded. Reference to Table 2 shows that the top track group 14 contains only one of the 201 top tracks. Consequently the number of particles recorded with a deflection of 14 C.U. is 200/201 of the correct amount. Using this method a series of correction factors may be calculated for deflections from 14 to 41 C.U.

(ii) *Failure of Counters in the Equipment.*—When the results are plotted as a deflection distribution it is found that, apart from the expected statistical scatter of the points about the smooth distribution, there is a marked tendency for the points corresponding to even values of D to be high and those corresponding to odd values of D to be low. This is due to counter failure in tray 3 during the experiment. The effect does not influence the overall shape of the distribution but introduces this undesirable spread of the points. This may be reduced, with a small sacrifice in resolution, by averaging the points in the following manner.

$$N_D^{AV}=\tfrac{1}{4}(2N_D+N_{D-1}+N_{D+1}),$$

where N_D=number of particles with deflection D.

(iii) *Accuracy of Deflection Measurement.*—In the previous analysis we have considered that the tracks were perfectly defined, i.e. they passed through the centre points of the C.U., and that the only source of deflection was the magnetic field. Errors in the determination of the deflection may arise from the following two sources.

(A) The spread due to the finite size of the counting units. Consider any two C.U. which constitute a top track. With a zero magnetic field and assuming no scattering, the intensity distribution in tray 3 of particles traversing these C.U. would have a triangular form (Rathgeber 1938). This distribution $f(D)dD$ (Fig. 9 (A)) falls to zero at $D=\pm 2$ C.U.

(B) The spread due to multiple scattering in the counter walls and in the air between the counter trays. A beam of mesons, homogeneous in momentum, passing through infinitely small C.U. in trays 1 and 2 would have the scattering distribution $g(D, p)dD$ in tray 3. This distribution is of Gaussian form for mesons homogeneous in momentum, the variance of the distribution depending on momentum.

These two effects are superimposed in the equipment, and we may find the resulting distribution $h(D, p)dD$ for any momentum by integrating the scattering distribution over the geometrical distribution. $h(D, p)dD$ now gives a measure of the deflection accuracy at any particular momentum. Since the distribution is not Gaussian, a convenient measure of deflection accuracy is given by the quartile deviation $\triangle D$. Calculating $\triangle D$ for various values of the momentum gives

$p>2\times 10^9$ eV./c., $\triangle D=0\cdot 7$ C.U. : $p=10^9$ eV./c., $\triangle D=1\cdot 0$ C.U.,

$p=5\times 10^8$ eV./c., $\triangle D=1\cdot 7$ C.U. : $p=2\cdot 4\times 10^8$ eV./c., $\triangle D=3\cdot 2$ C.U.

The calculation shows that for $p > 2 \times 10^9$ eV./c. the scattering is negligible compared with the geometrical spread. At the limit of resolution of the equipment, i.e. $D=1$ C.U. with $B=13{,}500$ gauss, the deflection error is $D=\pm 0\cdot 7$ C.U.

The expected zero field distribution may be obtained by numerically integrating $h(D, p)dD$ over the momentum spectrum of the incident mesons. This may be compared with that obtained by experiment.

IV. DISCUSSION OF RESULTS

(*a*) *Deflection Distributions*

A total of 6,056 single tracks which penetrated the 10 cm. of lead has been obtained for four values of the magnetic induction. These are plotted as deflection distributions together with a typical zero field distribution in Figure 9.

As a periodic check on the equipment, the magnet is demagnetized and a set of results taken in zero field. Figure 9 (A) shows one such zero field distribution. Due to the limited number of deflection units involved the deflection scale has been extended and the points plotted directly without the averaging process discussed previously. The full line and broken line curves show, respectively, the calculated zero field distributions including and neglecting multiple scattering, both normalized such that their areas agree with the number of particles recorded. The influence of the finite size of the counters on the calculated distribution has been neglected since it is smaller than the effect of the scattering.

The general agreement with the calculated distributions including scattering is good. This agreement confirms the assumptions made for the scattering calculation and shows that no other factors contribute to the spread of the zero field distribution. This zero field distribution also serves to check the alignment of the three resolving counter trays. For the distribution to be symmetrical it is necessary that these trays be no more than $0\cdot 5$ mm. out of line. Once this setting is obtained it is checked daily by means of an optical sight permanently attached to the equipment.

The deflection distributions obtained at four values of the induction are shown by the histograms B, C, D, E in Figure 9. These are obtained from approximately equal numbers of tracks for both directions of the magnetic field, and have been corrected and averaged in the manner described in the preceding section. The distributions for positive and negative particles are shown separately; those particles for which no deflection was detected are assumed to have a $+/-$ ratio of $1\cdot 25$. Each of the distributions is labelled with the value of the magnetic induction, the actual number of particles recorded, and the range of momenta calculated from the relation $p=3\cdot 9B/D$ MeV./c.

A study of the relation above shows that the deflection recorded in this equipment is the exact counterpart of the curvature of a cloud chamber track. Consequently, apart from the units involved, the deflection distributions may be compared directly with the specific curvature spectrum obtained by Blackett (1937). One requirement emphasized by Blackett is that, apart from the limitation imposed by the resolution of the equipment, the distribution must fall to zero as the deflection approaches zero for the total energy to be finite. This is equivalent to stating that above a certain momentum the momentum spectrum must have a negative exponent greater than two. This behaviour was not apparent to any appreciable extent in Blackett's curvature spectrum probably

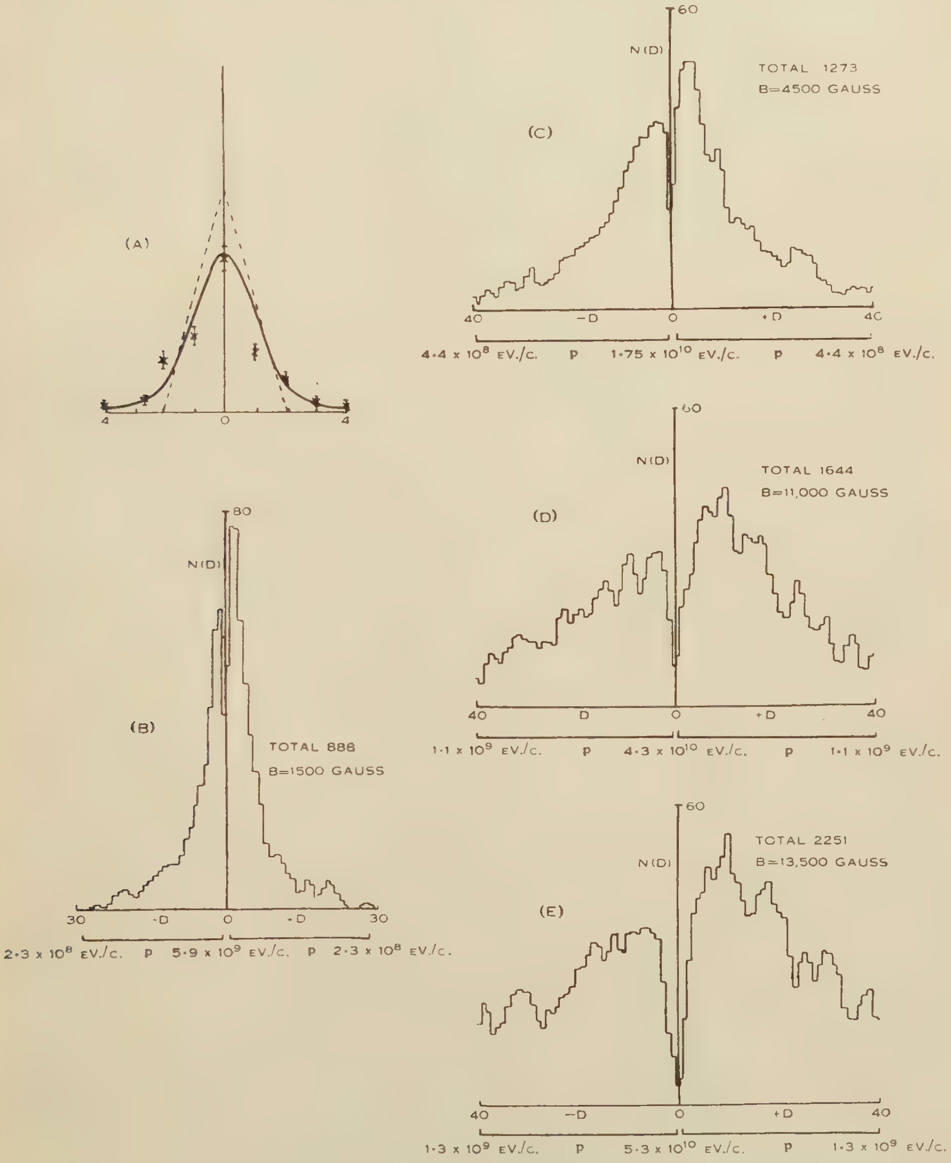


Fig. 9.—Deflection distributions obtained for various values of the magnetic induction.

because of insufficient resolving power; this will be discussed in more detail in a following section. However, in the distributions drawn in Figure 9, it is seen that as *B* is increased, corresponding to an increase in resolving power, this condition is more nearly fulfilled. For the highest field, of the 2250 particles recorded, only 9 tracks were found in the zero deflection unit.

The high field distribution ($B=13{,}500$ gauss) is made up of two sets of results taken over different periods. Both of these sets showed the same general shape as distribution (E) in Figure 9. It is interesting to note that in one of these distributions the anomaly observed by Blackett in the region of $2\cdot5\times10^9$ eV./c., i.e. a sharp drop in intensity, was quite marked. The other distribution, however, showed no real evidence for its existence, and when the results are combined the magnitude of the effect is only of the same order as other fluctuations in intensity. Consequently it is felt that, in view of the limited statistical accuracy at the present stage of the investigation, not much significance may be attached to this agreement.

(b) *Momentum Distribution*

The deflection distributions determine the momentum spectrum over overlapping ranges of momentum, the total range covered being from $5\cdot3\times10^{10}$ eV./c., corresponding to a deflection of 1 unit at 13,500 gauss, down to the cutoff (Montgomery 1949) imposed at $2\cdot4\times10^8$ eV./c. by the 10 cm. of lead. This cutoff is observed in the distribution taken at 1500 gauss and occurs for a deflection of 25 units. Two particles were recorded as having deflections of 26 and 28 units respectively, which were greater than this limit. This is to be expected, however, since the deflection accuracy in this region is determined mainly by scattering and is of the order of ±3 units.

The momentum distribution may be constructed by either of two methods.

(i) Directly from the deflection distribution.

$$\text{Since } p=\frac{3\cdot9B}{D}\ \text{MeV./c.}$$

and

$$F(p)dp=G(D)dD,$$

$$F(p)=\frac{D^2}{3\cdot9B}\cdot G(D). \qquad\qquad (4)$$

$G(D)$ is the smooth curve drawn through the deflection histogram and $F(p)$ then represents the differential momentum spectrum, i.e. intensity per MeV./c. momentum interval.

(ii) The corrected number of particles from each run may be grouped into momentum intervals of the desired size and plotted directly. The latter method is chosen in this paper as being more straightforward and less liable to statistical inaccuracy; also it is easier to assess the standard error of the points.

The positive and negative particles are combined and the data are plotted in Figure 10. The results obtained at different values of B are plotted separately and normalized to have the same number of particles with $p>2\times10^9$ eV./c. The figure becomes somewhat overcrowded if the standard error is shown for every point so the policy has been adopted of plotting the standard errors of only the end and mid-points of each run. A correction is necessary only for the higher deflection points ($D>14$ C.U.) in each run, thus the good agreement between corrected and uncorrected points from different runs justifies the method used to correct for magnetic cutoff.

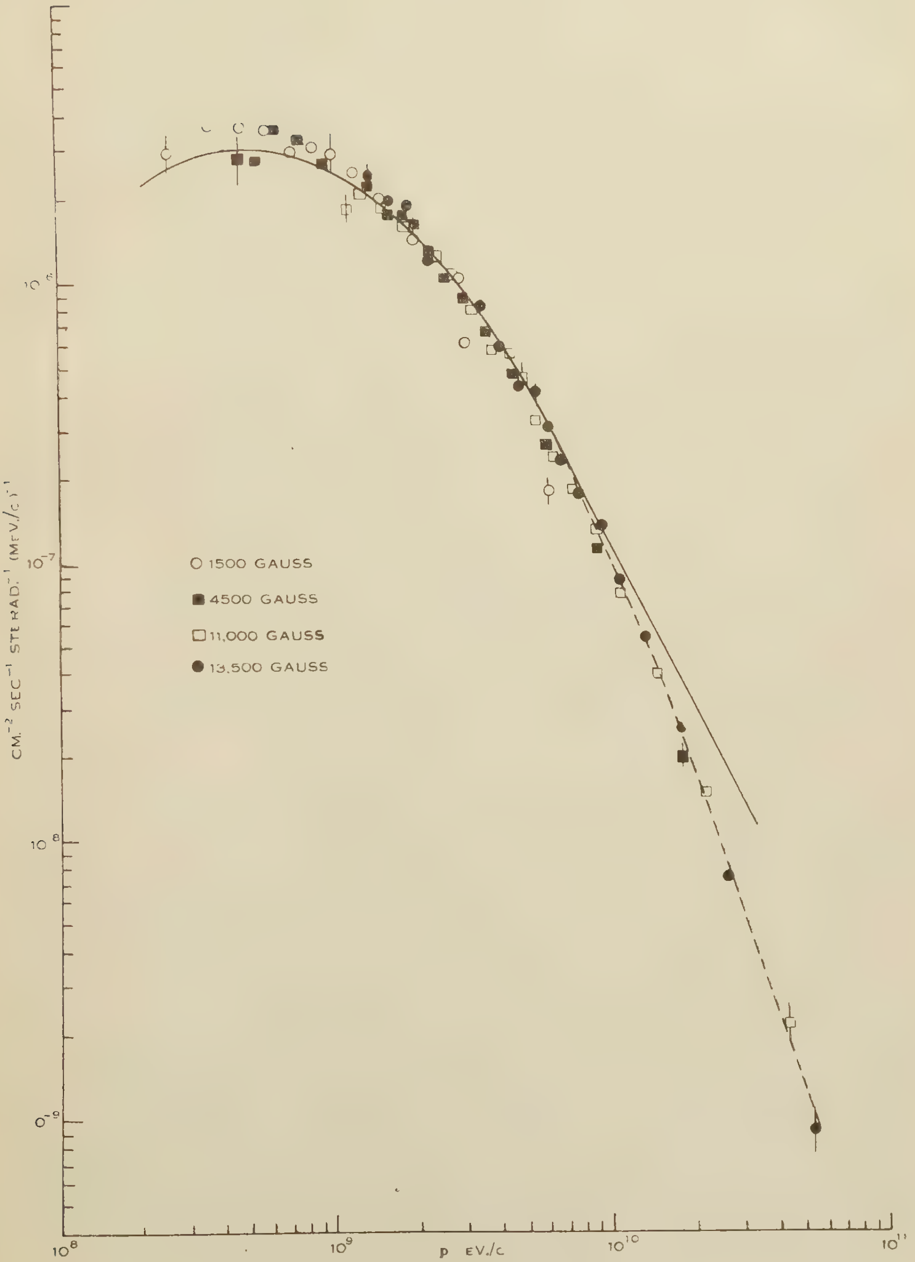


Fig. 10.—Differential momentum distribution of mesons at sea-level.

As a basis for comparison we have reproduced the momentum spectrum given by Rossi (1948) and adjusted our results to agree with it at $p=2\times10^9$ eV./c. This spectrum is derived from a consideration of both the cloud chamber and absorption measurements. The results agree with this curve over the region, 10^9 to 10^{10} eV./c., in which the cloud chamber results are most reliable. At the low momentum end of the spectrum the results confirm Rossi's conclusion that Wilson's (1946) data, which are the best available cloud chamber measurements in this region, are insufficiently corrected for magnetic cutoff. This indicates that the peak of the distribution at sea-level is in the region of 4×10^8 eV./c., a result which supports the absorption determinations. The points are consistently higher than the curve around the peak but since both measurement are liable to error in this part of the spectrum the agreement is considered satisfactory. It should be pointed out that even if the statistical accuracy was considerably higher, any irregularities in this region (Rogozinski and Lesage 1949) should be smoothed out due to scattering. However, at the present stage of the experiment, the influence of scattering on the general shape of the distribution is negligible compared to the statistical accuracy of the points.

Towards a high momentum limit the points begin to fall away more rapidly than shown by Rossi's curve until, above 10^{10} eV./c., the deviation becomes significant. The results lie quite smoothly on a curve (shown as a broken line in Fig. 10) which falls off appreciably faster with momentum than any of the previous results suggest. Since this deviation occurs at the limit of resolution of the cloud chamber determinations it is rather difficult to compare the results in this region. However, an estimate may be made by normalizing all the results to have an equal number of particles with $p>2\times10^9$ eV./c. and comparing the percentage which are recorded with $p>6\times10^9$ eV./c. and $p>10^{10}$ eV./c. The results of Blackett (1937), Jones (1939), Hughes (1940), and Wilson (1946) have been treated in this manner and compared with the figure obtained in this experiment. For the percentage of particles with $p>6\times10^9$ eV./c. all the measurements agree with the exception of Jones whose results give rather a high value. However, for $p>10^{10}$ eV./c. the percentages are Wilson 23 per cent., Blackett 21 per cent., Jones 32 per cent., and Hughes 26 per cent., while our experiment yields 16 per cent.

It is suggested that the reason that other observers have not recorded a fraction as low as 16 per cent. is the poor resolution in this region. This is supported by the fact that for the results obtained with $B=1500$ gauss, which is equivalent to reducing the resolution by a factor 9, the fraction of particles with $p>10^{10}$ is 23 per cent. This corresponds to a limit of resolution, i.e. a 100 per cent. uncertainty in deflection, of 8×10^9 eV./c. On this basis it appears that while Blackett and Wilson have achieved slightly higher resolution than this figure, the resolving power realized in the experiments of Hughes and Jones is not as high.

It may be concluded that the resolution as deduced from tracks photographed in zero field is better than that actually achieved when operating in a magnetic field, possibly due to less favourable thermal conditions. Additional evidence favouring this conclusion is found in the work by Blackett (1937).

On the basis of our results it is possible to form an estimate of the specific curvature spectrum which could be expected from Blackett's work. This is done by using the high field deflection distribution, averaging to give intervals corresponding to $d\sigma = 0 \cdot 1$ cm.$^{-1}$ and analytically introducing the spread expected from Blackett's zero field distribution. The result showed a drop in intensity for the first two intervals of σ of the same order as in Figure 9 (C) for which a resolution of $1 \cdot 75 \times 10^{10}$ eV./c. is claimed. The very small drop recorded by Blackett indicates that the resolving power realized during operation is somewhat smaller than the estimated value.

(*c*) *The Excess of Positive Particles*

Inspection of the deflection spectra (Fig. 9) shows that the number of positive particles which pass through 10 cm. of lead exceeds that of the negative particles. The trend of this excess in terms of momentum was obtained by averaging the uncorrected measurements over momentum intervals separately for positive and negative particles, and forming their ratio. The range of each momentum interval and the standard error calculated from the number of particles in each interval are shown by the breadth and the height, respectively, of the rectangles drawn in Figure 11 (A).

The geometry of the Geiger counters causes particles which have not been deflected by the field to appear not only in the zero deflection unit but also in the next two C.U. on either side (Fig. 9 (A)). The contribution to the second C.U. is small. Therefore only the deflection units +1 and −1 have been omitted from the calculations of the ratios for fields of 1500 and 4500 gauss. This omission is not necessary for the high fields as then the number of undeflected particles is appreciably smaller than the number of particles in the first deflection units.

To decrease the statistical errors greater momentum intervals than in a previous note (Caro, Parry, and Rathgeber 1950) have been chosen. Comparing the results for fields of 1500, 4500, and 13,500 gauss it is found that in fact the scatter of the points has been reduced without affecting the general trend. However, the new measurement at 11,000 gauss, which has been made since the writing of the above-mentioned note, affects our conclusions at the higher momenta.

To reduce further the statistical errors we have formed the means of the measurements at the several fields for momenta of 0·6, 1·5, 4, and 15 BeV./c. They are shown together with the results of other measurements in Figure 11 (B). The straight line obtained in our previous note which is traced on the same figure gives a satisfactory first approximation to the trend of the measurements. The observation that the point at 4 BeV./c. misses the line by more than twice its standard error and that our results as well as most of the others are consistently high in that region make us prefer the curve as a better approximation.

No satisfactory theory of the excess of positive particles at sea-level with which the above results could be compared has as yet been worked out. A calculation (Bassi *et al.* 1949*b*) has been based on the assumption that an excess of positive particles is created by a primary cosmic ray particle in its first nuclear collision, and that the subsequent collisions which create equal numbers of positive and negative particles occur at an appreciably lower height. These authors have shown that their calculations can be made to agree by adjustment of several parameters with a curve which they derive from their own and Wilson and Owen's (1949) measurements. This latter curve is similar to that obtained in our experiment. Since, according to the evidence found in photographic emulsions, a primary particle loses most of its energy in a collision with a single nucleus, whether the process be multiple or plural, the above theory has to be treated with reserve. Rough theoretical considerations (Lewis, Oppenheimer, and Wouthuysen 1948) make us expect that the excess of positive particles decreases as the inverse of the square root of the energy. Our curve indicates that such a decrease might set in at momenta of about 5 BeV./c. The decrease towards lower momenta could be due to secondary effects. Thus the excess of positive particles at momenta greater than 5 BeV./c. is of considerable interest

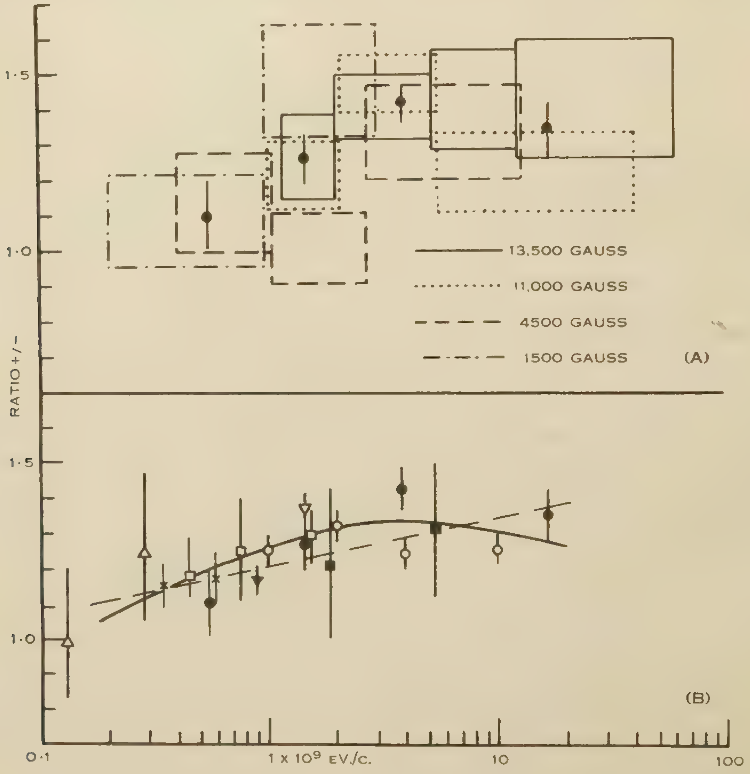
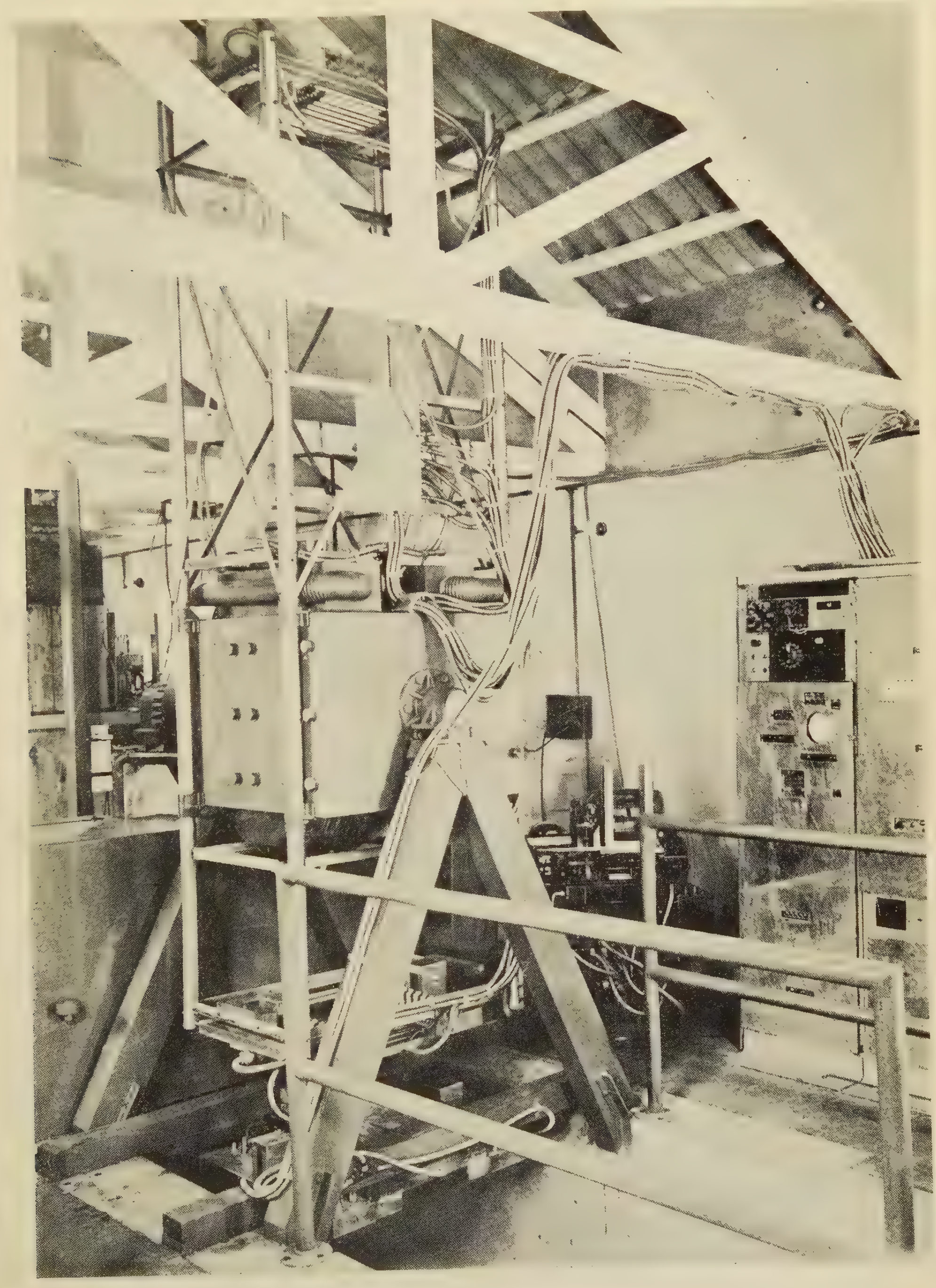


Fig. 11 (A).—Ratio of positive to negative mesons measured with the cosmic ray spectrometer.

Fig. 11 (B).—Comparison of measurements of the ratio of positive to negative mesons.

△ Conversi (1949). × Quercia, Rispoli, and Sciuti (1948). □ Bassi *et al.* (1949*a*). ▼ Nereson (1948). ○ Owen and Wilson (1949). ▽ Brode (1949). ■ Leprince-Ringuet and Crussard (1937). ● Averages of Fig. 11 (A).

COSMIC RAY MESONS AT SEA-LEVEL



General view of the cosmic ray spectrometer.

for the theory of nuclear forces. Our measurements at these energies will be resumed with increased resolving power as soon as the present investigations are completed.

V. Conclusion

The momentum spectrum of C.R. mesons has been measured by a magnetic deflection method up to a momentum of 5×10^{10} eV./c. Over the range $2\cdot4\times10^{8}$ eV./c. to 10^{10} eV./c. the results are in agreement with previous determinations by cloud chamber and absorption methods. At momenta above 10^{10} eV./c., however, the spectrum falls off faster than previous experiments suggest. The integral spectrum in this region may be represented by a power law with an exponent of -2 which is approximately the value obtained from absorption methods at depths of 50 to 250 metres of water.

The ratio of positive to negative particles has been calculated over the same momentum range. Contrary to theoretical expectation, this ratio increases with increasing momentum up to 5×10^{9} eV./c. For higher momenta the $+/-$ ratio levels off and perhaps decreases at the limit of resolution of the equipment.

VI. Acknowledgments

The authors wish to thank Professor L. H. Martin for his interest and encouragement throughout the work, to Mr. K. Nicholson, Mr. P. Goodman, and Mr. J. Swift for their help in the construction and operation of the equipment, and to Kalamazoo Pty. Ltd., for their assistance with the Powers machine. The construction of the equipment was made possible by a grant from C.S.I.R.O.

VII. References

Bassi, P., Clementel, E., Filosofo, I., and Puppi, G. (1949*a*).—*Phys. Rev.* **76** : 854.

Bassi, P., Clementel, E., Filosofo, I., and Puppi, G. (1949*b*).—*Nuovo Cim.* **6** : 484.

Blackett, P. M. S. (1937).—*Proc. Roy. Soc.* A **159** : 1.

Brode, R. B. (1949).—*Phys. Rev.* **76** : 468.

Caro, D. E., and Parry, J. K. (1949).—*J. Sci. Instrum.* **26** : 374.

Caro, D. E., Parry, J. K., and Rathgeber, H. D. (1950).—*Nature* **165** : 688.

Coggeshall, N. D. (1947).—*J. Appl. Phys.* **18** : 855.

Cork, J. M., and Wadey, W. G. (1942).—*Rev. Sci. Instrum.* **13** : 369.

Conversi, M. (1949).—*Phys. Rev.* **76** : 311.

Curtiss, L. F. (1932).—*Bur. Stand. J. Res. Wash.* **9** : 815.

Hughes, D. J. (1940).—*Phys. Rev.* **57** : 592.

Jones, H. (1939).—*Rev. Mod. Phys.* **11** : 235.

Leprince-Ringuet, L., and Crussard, J. (1937).—*J. Phys. Radium* **8** : 207.

Lewis, H. W., Oppenheimer, J. R., and Wouthuysen, S. A. (1948).—*Phys. Rev.* **73** : 127.

Montgomery, D. J. X. (1949).—"Cosmic Ray Particles." p. 354. (Princeton Univ. Press.)

Nereson, N. (1948).—*Phys. Rev.* **73** : 565.

Owen, B. G., and Wilson, J. G. (1949).—*Proc. Phys. Soc.* A **62** : 601.

Quercia, I. F., Rispoli, B., and Sciuti, S. (1948).—*Phys. Rev.* **73** : 516.

Rathgeber, H. D. (1938).—*Z. Phys.* **109** : 273.

Regener, V. H. (1943).—*Phys. Rev.* **64** : 250.

Rogozinski, A., and Lesage, M. (1949).—"Cosmic Radiation." Colston Pap. (Butterworth's Scientific Publications : London.)

Rossi, B. (1948).—*Rev. Mod. Phys.* **20** : 543.

Swann, W. F. G. (1940).—*J. Franklin Inst.* **229** : 300.

Wilson, J. G. (1946).—*Nature* **158** : 414.

OBSERVATIONS OF THE SPECTRUM OF HIGH-INTENSITY SOLAR RADIATION AT METRE WAVELENGTHS

IV. ENHANCED RADIATION

By J. P. Wild*

[*Manuscript received September* 11, 1950]

Summary

Observations are described of the spectrum of " enhanced radiation " from the Sun (i.e. the radio-frequency radiation which maintains a high but variable level for periods of hours' or days' duration) in the frequency range 70–130 Mc/s. This radiation is known to be received from the direction of sunspots and to show circular polarization. For the purpose of presenting results, two components are recognized, viz. a background continuum which varies gradually with time and frequency, and short-lived, narrow-band bursts (" storm bursts "). The behaviour of the two components, and the relation between them during periods of high level (" noise storms ") are described. A detailed analysis is given of the properties of recorded storm bursts. The distribution of recorded bursts with frequency was found to be markedly non-uniform, e.g. a pronounced minimum at 89 Mc/s. was present. The possibility that the background continuum is due to the resultant of a large number of bursts is discussed.

I. Introduction

This is the final paper of a series describing some observations of the radio spectrum of high-intensity solar radiation in the frequency range 70–130 Mc/s. (Wild and McCready 1950 ; Wild 1950*a*, 1950*b*). The present paper gives details of the radiation, associated with sunspots, which maintains a high but variable level for periods of hours or days at a time. Characteristics of this radiation have been described by a number of observers (see, for instance, Appleton and Hey 1946 ; Allen 1947 ; McCready, Pawsey, and Payne-Scott 1947 ; Ryle and Vonberg 1948 ; Pawsey 1950) working at single frequencies between about 50 and 200 Mc/s. These are summarized below :

(i) The radiation at any one frequency appears as a steady or slowly-varying background level with frequent short-lived bursts.

(ii) The radiation is circularly polarized. At times of high intensity, the degree of polarization is usually nearly complete. The polarization of the short-lived bursts is usually of the same sense as that of the background radiation.

(iii) The radiation is received from the direction of sunspots and originates in an area of the Sun of the same order of size as a sunspot. The radiation has been reported to reach a sharp maximum near the central meridian passage of a large spot group.

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

This radiation constitutes by far the greater part of high-intensity solar radiation observed at metre wavelengths. In conformity with previous writers, we shall refer to it as " enhanced radiation " (Pawsey 1950), and to its periods of occurrence as " noise storms " (Allen 1947).

II. OBSERVATIONS OF THE SPECTRUM DURING NOISE STORMS

(*a*) " *Background Continuum* " *and* " *Storm Bursts* "

Between February and June 1949 the solar spectrum in the frequency range 70 to 130 Mc/s. was observed for many hours during noise storms. The apparatus has already been described (Wild and McCready 1950). The spectrum was displayed on a cathode-ray tube showing receiver output (vertically) and frequency (horizontally). Continuous photographic records were taken during periods of interest.

When severe noise storms were in progress, the spectrum showed a high steady level which usually extended over the whole frequency range, rarely showing any steep variations of intensity with frequency.

Above this " background continuum " short-lived bursts occurred, sometimes many per minute. Some sample portions of the record showing these bursts are reproduced in Plate 1. The spectrum of the bursts nearly always showed two features. Firstly, the spectrum at any instant was narrow, typically about 4 Mc/s. between points of quarter-maximum intensity. (This narrow bandwidth is consistent with reports by Ryle and Vonberg (1948) and Payne-Scott, Yabsley, and Bolton (1947) of a lack of coincidences between bursts observed at different frequencies.) Secondly, the mid-frequency of a burst remained approximately constant throughout its lifetime. The shape of the spectral profile varied from burst to burst, although perhaps most were approximately symmetrical with maximum intensity near the mid-frequency. Occasionally the spectral profile at a particular time showed two or more maxima (see Plate 1 (*c*)) ; in such cases it was usually clear from a study of the temporal variations of each maximum that two or more bursts at neighbouring frequencies were merged together. The simplest of the bursts rose steadily with time to their maximum amplitude and then decayed steadily ; it was common, however, for a burst to rise and fall several times, with no appreciable change of mid-frequency, before finally subsiding. Sometimes the rise and fall of bursts were extremely rapid—much more rapid than those observed in " outbursts " and " isolated bursts " ; occasionally the intensity changed by a factor of the order of a hundred in a period of one-third of a second.

These narrow-band bursts which are a feature of noise storms will be referred to as *storm bursts*.*

Apart from storm bursts, variations in the background continuum were usually too small or too slow to be readily detected by the eye when watching the display. At times of very great burst activity, however, the continuum often showed irregular " ripples " over a wide frequency range. Furthermore, on two occasions which are discussed in Section II (*e*), the continuum reached an exceptionally high level for several minutes, fluctuating more rapidly than usual. Such phenomena were uncommon, and for the purpose of presenting data we shall regard the enhanced radiation as consisting of two discrete components—the relatively steady background continuum, and the narrow-band storm bursts. Since these components differed markedly in their spectral and temporal characteristics, a clear distinction could be made between them.

---

* The spectrum of storm bursts was designated in Part I (Wild and McCready 1950) as " Type I ".

Storm bursts, which we have defined in terms of spectral characteristics, were sometimes observed on days when no detectable background continuum was present, and occasionally as isolated occurrences on " quiet " days. However, as indicated in Figure 1, there was a high correlation between the rate of occurrence of storm bursts on a given day and the average level of the background continuum.

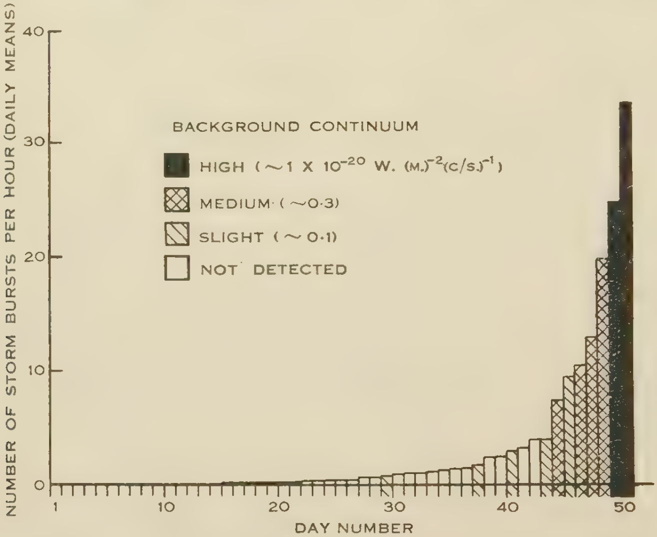


Fig. 1.—Diagram indicating the correlation between the average number of storm bursts per hour and the average intensity of the background continuum within the observed frequency range (70 to 130 Mc/s.) during 51 days of observation. Each day's observation is represented by a vertical strip. The average number of storm bursts per hour is given by the height of the strip, and the average background continuum by the shading of the strip. The strips are arranged in order of increasing height.

(*b*) *Presentation of Data on Storm Bursts*

In presenting quantitative data on storm bursts, it is convenient to divide the bursts into several groups according to size. The following three size-groups have been adopted:

Group 1: maximum intensity between 1 and 3 units of 10^{-20} W. m.$^{-2}$ (c/s.)$^{-1}$.*

---

* As in previous papers, specified intensities refer to one plane-polarized component of the radiation.

Group 2 : maximum intensity between 3 and 9 units.

Group 3 : maximum intensity greater than 9 units.

Bursts with maximum intensity less than 1 unit have been disregarded.

It is also convenient to introduce a quantity that indicates the frequency distribution of radiation emitted in storm bursts during a given period. The most desirable quantity would be the mean intensity at each particular frequency. In many cases, however, the records were not complete enough for this quantity to be evaluated. Instead we evaluate a " burst index ", $B(f)$, for a given period, defined by

$$B(f)=\frac{1}{T}\sum_r A_r(f),$$

where T denotes the duration of the period,

$A_r(f)$ denotes the maximum amplitude reached, at frequency f, by an individual burst (distinguished by the suffix r-), and

$\sum_r$ denotes summation over all bursts (of group 1 or larger) occurring within the period.

When a large number of bursts is considered, the burst index probably gives a fair indication of the mean-intensity distribution, since, as will be shown, the lifetime of bursts in the observed range does not seem to depend upon frequency.

(c) *Solar Activity and Noise Storms during the Period of Observation*

The principal solar events during the period of observation were associated with a sunspot group which crossed the central meridian on May 11·5, 1949, in latitude 17 °S. (see Newton 1949). The spot group produced several flares as it approached the meridian. The most severe noise storm recorded appeared to be associated with this spot group. The storm was first observed on May 9 and continued until May 13 ; it reached a maximum on May 10, when storm bursts were more numerous and the background continuum higher than on any other day throughout the observing period.

The second most active day was April 14. The noise storm on this day occurred 27·5 days before the central meridian passage of the spot group cited above, and may possibly have been due to the same spot group on its previous revolution (then not visible). However, a number of fairly large spot groups were visible at this time and the origin of the noise storm is therefore uncertain.

A detailed description of the enhanced radiation received during the two active days April 14 and May 10 is given in Figures 2 and 3. The following data are plotted :

(i) The time, mid-frequency and amplitude, in the three groups, of all storm bursts recorded during the periods. (The recorded bursts represent a large majority and fair sample of all the bursts that occurred except during periods of very high activity, when larger bursts were favoured.)

(ii) The burst index over 20-minute periods.

(iii) The spectrum of the background continuum at 20-minute intervals.

(iv) The burst index and average spectrum of the background continuum for each day.

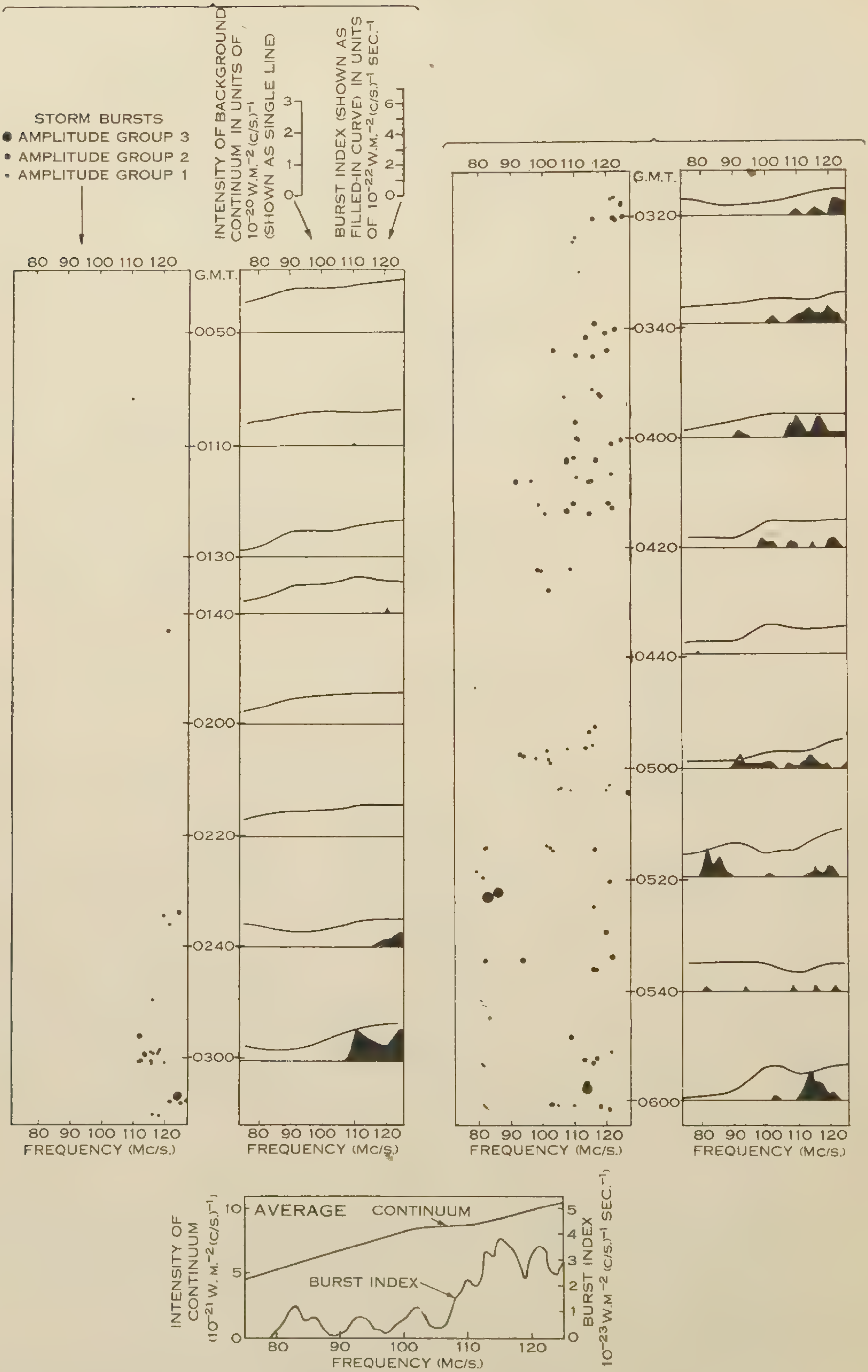


Fig. 2.—Description of enhanced radiation recorded on April 14, 1949. The left-hand columns show the time of occurrence, mid-frequency and amplitude (in three groups) of storm bursts. The right-hand columns show the spectrum of the background continuum at 20-minute intervals (single lines) and the "burst index" over 20-minute periods (filled-in curves).

Some features of noise storms suggested by these records and similar records of other active days are summarized below.

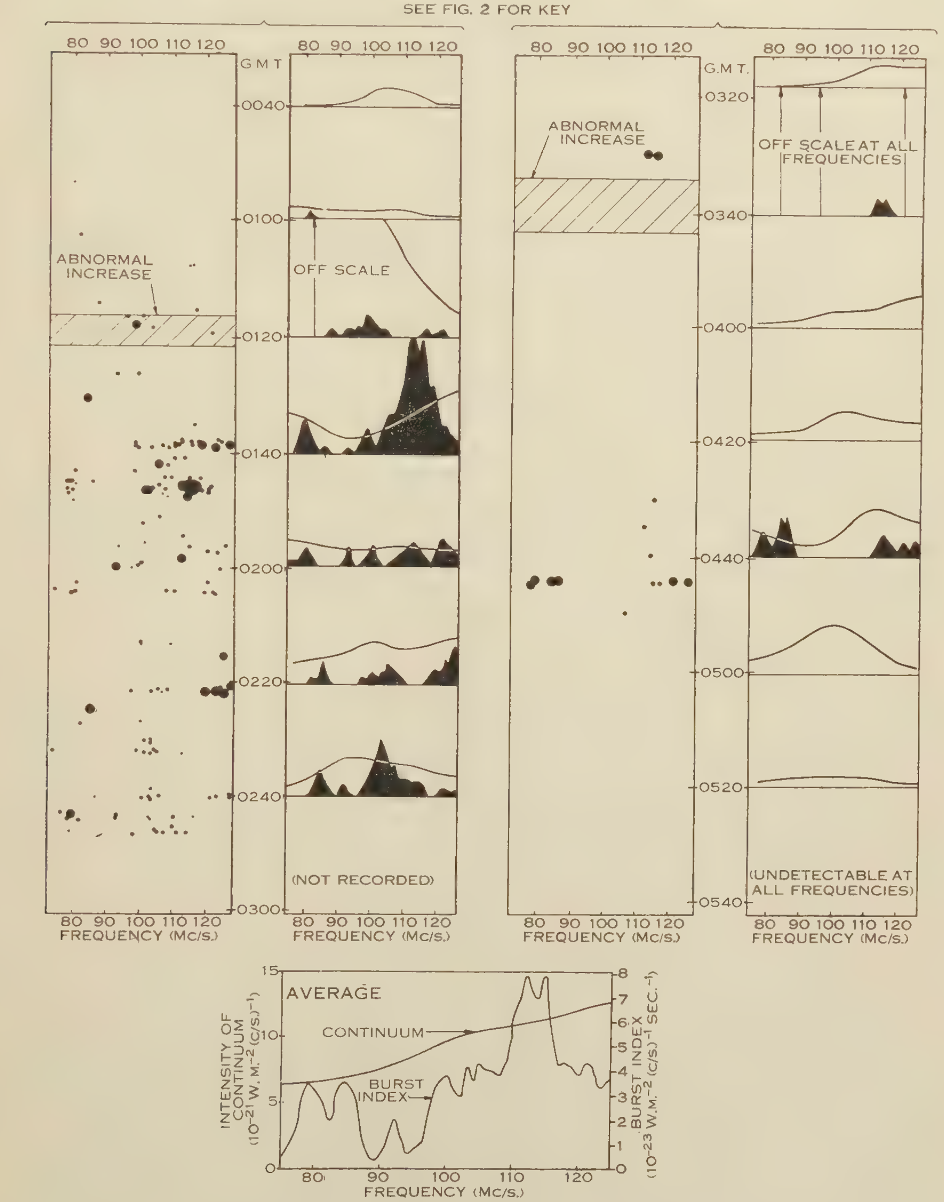


Fig. 3.—Same data as in Figure 2 for May 10, 1949. The two "abnormal increases" each accompanied a solar flare.

(*d*) *Summarized Features of Noise Storms*

(i) The spectrum of the background continuum appeared to change continuously with time. Although this change was not usually detectable by the eye when watching the cathode-ray-tube display, the spectrum commonly showed marked changes in profile in periods of 20 minutes. The "spot frequency" spectra given by Hey, Parsons, and Phillips (1948) suggest a similar effect.

(ii) The average intensity of the background continuum on most of the recorded days increased steadily with frequency in the observed range. On the other hand, Hey, Parsons, and Phillips (1948) have reported that the maximum intensity of the "continuous emission from sunspots" often appears to be in the region of 60 Mc/s.

(iii) The spectrum due to the integrated effect of storm bursts (as indicated by the burst index) over 20-minute periods shows no apparent relation to that of the background continuum.

(iv) A background continuum of high level can sometimes be present for a period of an hour or more with no appreciable storm-burst activity at any frequency in the observed range (see, for instance, April 14, 00 hr. 50 min. to 02 hr. 30 min.).

(v) Storm bursts showed a marked tendency to occur during brief active periods of about 1 or 2 minutes' duration. Relatively "quiet" intervals between these active periods varied from a few minutes to several hours, perhaps most commonly between 10 and 30 minutes. (This effect was more marked on records of some of the less stormy days than on the two records reproduced here.)

(vi) On some days storm bursts seemed to occur in restricted frequency bands. Thus on May 9, 87 per cent. of observed bursts occurred in the lower half of the observed range (72 to 100 Mc/s.), whereas on June 27, 100 per cent. occurred in the upper half (100 to 128 Mc/s.). On April 14 (Fig. 2), from 02 hr. 30 min. onwards, bursts occurred at first at the higher frequencies only, but the low frequency limit of burst activity appeared to decrease gradually through the day. On all days there appeared to be a scarcity of bursts at about 88 to 90 Mc/s. This anomaly will be examined in Section III.

(*e*) *Abnormal Increases in Level at the Time of Solar Flares*

The features described above indicate the normal behaviour of the spectrum of enhanced radiation. Two exceptional events, somewhat similar to one another, occurred during the severe noise storm of May 10 and each coincided with a solar flare. On each occasion the background continuum rose steadily to a value many times its former level in a period of a few minutes. At times the level showed short-lived increases to exceptionally high values; unlike the characteristic storm bursts, these short-lived increases were of broad bandwidth (tens of megacycles per second). Some 5 to 15 minutes after the start, the level gradually subsided to its normal level.

The intensity variation of the greater of the two events is shown in Figure 4 for frequencies 75, 100, and 125 Mc/s. A flare of importance 2, located above the active sunspot discussed in Section II (*c*), reached its maximum at 03 hr. 40 min. G.M.T., about the time of the maximum radio intensity. Both the onset and decay of the visual and radio phenomena also seemed to show good correspondence in time. The spectrum showed a large increase of intensity with frequency throughout the phenomenon.

The lesser of the two events, which started at 01 hr. 17 min., differed from the greater in two main features: firstly, the associated flare, which was rated as importance 1, was located (at 7 °S. 35 °W.) above a spot group far removed from the main active group (at 17 °S. 18 °E.); secondly, the spectrum showed a marked *decrease* of intensity with frequency.

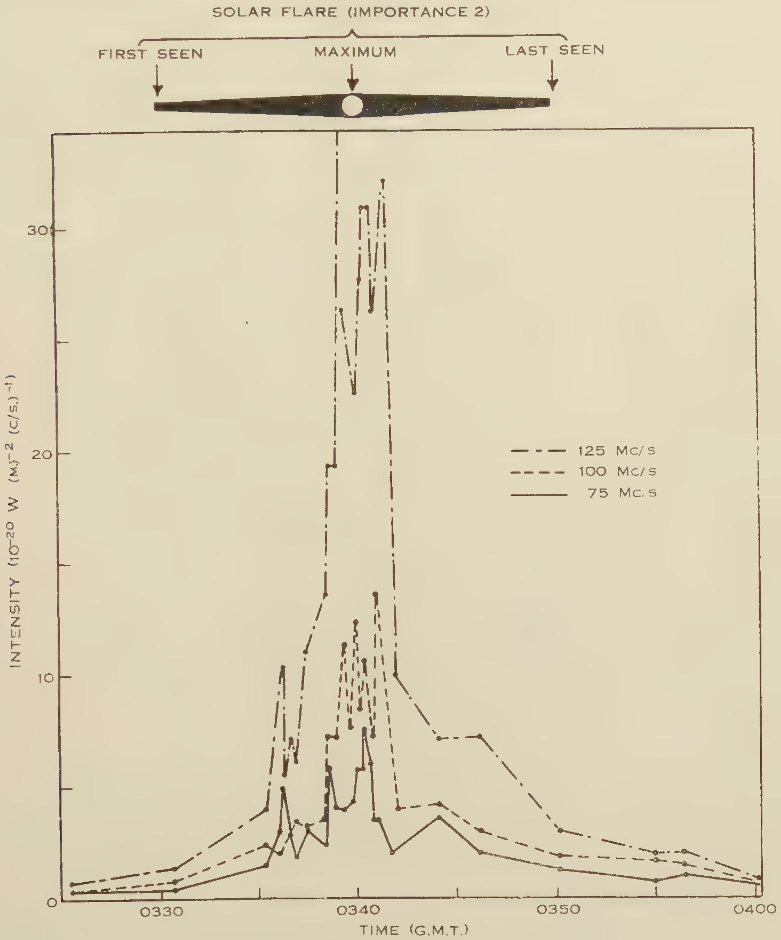


Fig. 4.—Plot of intensity at 75, 100, and 125 Mc/s. during an abnormal increase in level on May 10, 1949. (Data of the accompanying solar flare were supplied from the Commonwealth Observatory, Canberra.)

It should be remarked that the spectra of these abnormal increases showed no resemblance to the flare-accompanying " outbursts " described in Part II (Wild 1950*a*), and therefore were not considered under that category.

III. Analysis of Storm-Burst Data

We now proceed to give a detailed analysis of recorded storm bursts of amplitude-group 1 or more. These number about 670. The data have been derived entirely from photographic records.

We shall examine first the distribution of bursts with (*a*) frequency, (*b*) bandwidth, (*c*) lifetime, and (*d*) amplitude, and subsequently (*e*) the correlations between these four variates taken in pairs.

(*a*) *The Distribution of Bursts with Frequency*

The burst index for all recorded bursts is shown in Figure 5. The main features of this curve are a minimum at 89 Mc/s., a maximum on the low frequency side of this minimum with a peak at 79 Mc/s., and another maximum on the high frequency side with a peak at 113 to 115 Mc/s. In Figure 6 is shown the distribution of recorded bursts with mid-frequency. This histogram shows the same main features as Figure 5.

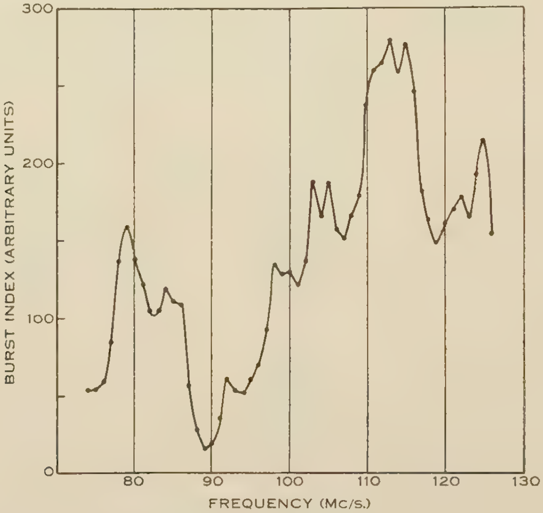


Fig. 5.—The variation with frequency of the burst index for the total period of observation.

The possibility that these features may be due to systematic errors in the calibration of intensity as a function of frequency has been rejected for the following reasons :

(i) No similar features have been found in the observed distribution of other types of high-intensity solar radiation using the same apparatus, e.g. the background continuum of enhanced radiation, and the integrated spectra of isolated bursts (see Part III (Wild 1950*b*)).

(ii) If we suppose that systematic errors were considerably greater than those estimated (see Part I (Wild and McCready 1950)), and were as much as three to one in intensity, and if we draw the smoothest possible curve on the basis that all such errors would so operate as to tend to eliminate the features, then we obtain the dotted curve of Figure 6. This curve shows the same main features as before (though of reduced magnitude), except that no peaky maximum near 115 Mc/s. is present.

It is important to inquire if these features are likely to reflect the shape of the real distribution (i.e. the distribution that would be given by an infinite number of observations over an infinitely long period). It has been calculated that if (i) each of the bursts in Figure 6 were independent of one another, and (ii) the real distribution were a smooth curve having neither maxima nor minima, then the probability of obtaining a random sample with dispersion as great as that of the observed sample would always be much less than 10^{-6}. We can therefore say that at least one of the suppositions (i) and (ii) is untrue. Now the observation that bursts tended to occur grouped in different frequency bands on different days is an indication that the bursts are not entirely independent events. For this reason it has not been possible to establish with certainty that the main features in the observed sample are also present in the real distribution. Observations of enhanced radiation on a large number of days would be required to decide this point.

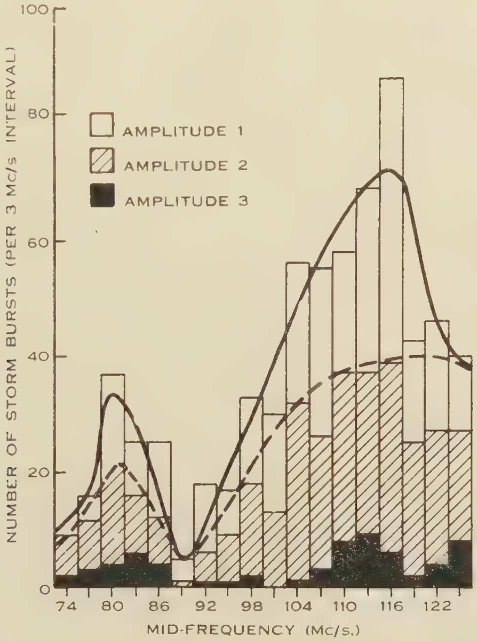


Fig. 6.—Histogram showing the number of storm bursts of amplitude groups 1, 2, and 3, recorded during the total period of observation, in successive 3 Mc/s. mid-frequency groups.

(*b*) *The Distribution of Bursts with Bandwidth*

The distribution of recorded bursts with bandwidth (measured between points of quarter-maximum intensity) is shown in Figure 7. The most common bandwidth is seen to be about 4 Mc/s.

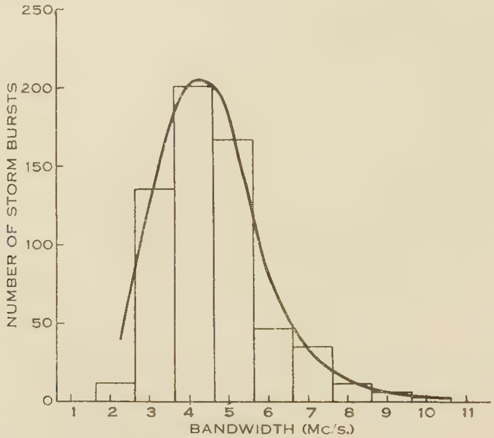


Fig. 7.—Histogram showing the distribution of recorded storm bursts with bandwidth.

The bandwidth of the receiver (0·3 Mc/s.) was considerably less than observed bandwidths, so that significant errors are not to be expected on this account.

(*c*) *The Distribution of Bursts with Lifetime*

The distribution of recorded bursts with lifetime (again measured between points of quarter-maximum intensity) is shown in Figure 8. Owing to the large

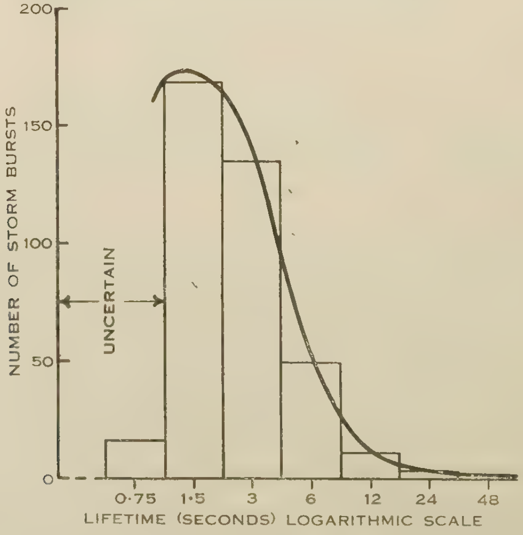


Fig. 8.—Histogram showing the distribution of recorded storm bursts with lifetime.

dispersion of lifetimes, it was found convenient to use a logarithmic scale of lifetime. Each interval of the histogram corresponds to a two-to-one range in lifetime. The number of recorded bursts suitable for lifetime measurement was rather less than the total number because the recording technique was such that the starts of bursts were sometimes missed (see Part I (Wild and McCready 1950)). For this reason also, a large proportion of very short bursts (less than one second, say) which were seen on the cathode-ray-tube display were not recorded. Thus the histogram is not reliable for small lifetimes.

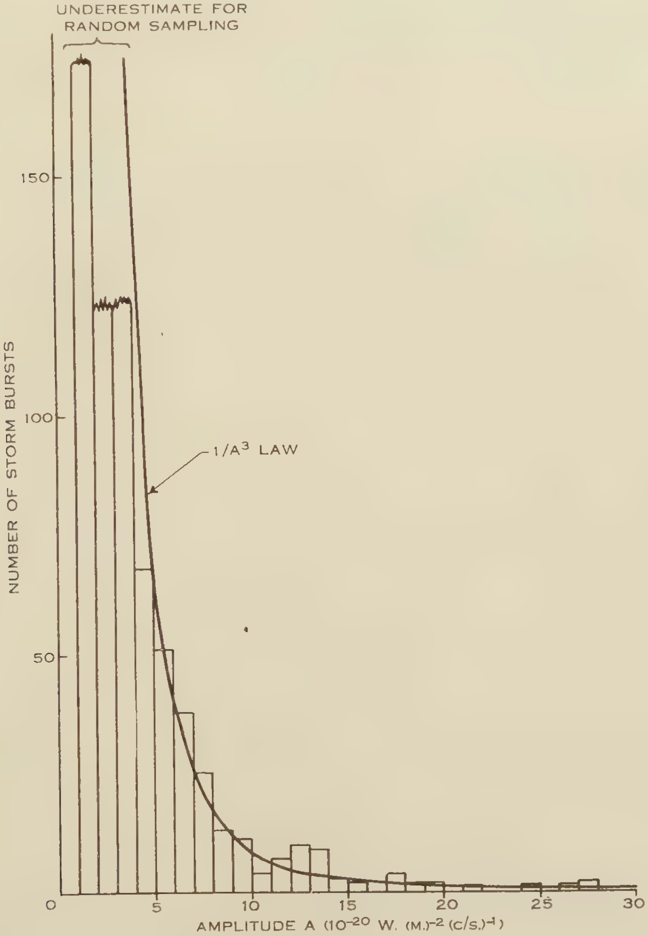


Fig. 9.—Histogram showing the distribution of recorded storm bursts with amplitude (i.e. maximum intensity at peak frequency).

Ryle and Vonberg (1948), observing on an apparatus with an output time constant of about half a second, at frequencies of 175 and 80 Mc/s., found that bursts of this kind usually lasted for 1 to 20 seconds. This result is seen to be in good agreement with the present data.

(*d*) *The Distribution of Bursts with Amplitude*

The distribution of recorded bursts with amplitude (i.e. maximum intensity) is shown in Figure 9. In the range of amplitudes from about 4 units upwards the data represent a fairly complete record of all bursts that occurred while observation was in progress. In this range the distribution was found to give a fairly good fit to the law $1/A^3$, where A denotes amplitude, as shown in the figure. However, the number of bursts of smaller amplitude in the histogram is likely to be considerably less than the total number that actually occurred because many small bursts were disregarded at times of very high activity.

(*e*) *Correlations*

The correlation between mid-frequency, bandwidth, lifetime, and amplitude of recorded bursts was examined. No definite correlation was found between any two of these four variates when taken in pairs. There was, however, slight evidence of a tendency for the average bandwidth to be greater at mid-frequencies above 90 Mc/s. than below, and for the average lifetime to increase with both bandwidth and amplitude.

IV. The Relation between the Background Continuum and Storm Bursts

In giving data on enhanced radiation we have recognized two components—the background continuum and the storm bursts. This distinction was made possible because any short-period fluctuations in the continuum other than the distinctive bursts were usually so small that they could scarcely be noticed when watching the spectrum display. The distinction may not, however, be a real one ; as pointed out by McCready, Pawsey, and Payne-Scott (1947) and Ryle and Vonberg (1948), the background level may be the resultant of a large number of bursts, only the larger ones being observed as distinct phenomena. This hypothesis is supported by (i) the observation (Ryle and Vonberg 1948) that the circular polarization of both bursts and background is nearly always of the same sense, and (ii) the good general daily correlation between the two components.

On the other hand, the opposite view—that the two components are fundamentally different—is suggested by (i) the usual small degree of fluctuation (other than recognizable bursts) in the background continuum, (ii) the lack of correspondence between the spectrum of the continuum and the integrated spectrum of the bursts, and (iii) the fact that during some periods a high continuum is observed with no appreciable burst activity, while at others bursts are observed with no detectable continuum. However, it appears possible to reconcile the former hypothesis with the latter three points by postulating a suitable amplitude distribution of the bursts (i.e. a suitable relationship between the amplitude, A, of individual bursts and the probability of a burst of amplitude between A and $A+dA$ occurring in unit time), the shape of the distribution being allowed to change slowly with time. It is clear, for instance, that if the amplitude distribution were such that most of the radiation was due to very small bursts (i.e. very small compared with the average level), then the aggregate effect would be a fairly steady background level. Thus the whole question depends critically upon the amplitude distribution, and in particular the population of the small-amplitude groups where direct measurement of individual bursts is not possible. The problem as to what kinds of amplitude distribution lead to the observed effects is beyond the scope of the present paper. It may be stated, however, that plausible distributions, containing adjustable parameters, have been found which are capable of describing, in terms of bursts alone, the wide variety of types of enhanced radiation observed (the " type " being determined by the values of the parameters). It is concluded, therefore, that the hypothesis is not unreasonable.

V. Conclusion

The enhanced radiation has been discussed theoretically by several authors (Martyn 1947 ; Giovanelli 1948 ; Ryle 1948 ; Bohm and Gross 1949 ; Haeff 1949 ; Bailey 1950 ; Bunemann 1950) and various mechanisms have been suggested to explain the generation of the observed high intensities. In general the theoretical predictions depend critically on quantities specifying the complex physical conditions that exist in the neighbourhood of active sunspots. In many cases we have little or no knowledge of such quantities. For this reason it has not so far been possible to decide in favour of any particular one of the suggested mechanisms.

The present observations suggest a possible new approach to the theoretical problem. The results indicate that storm bursts, which constitute part or all of the enhanced radiation, have fairly well defined properties (bandwidth, lifetime, etc.). Consideration of the physical significance of these particular phenomena may lead to a better understanding of the enhanced radiation.

Perhaps the most puzzling result of these observations is the marked lack of uniformity in the distribution of storm bursts with frequency during the 4-month period of observation, in particular the consistent scarcity of bursts in the vicinity of 89 Mc/s. It seems desirable to find out whether or not this effect is due to inadequate sampling before discussing its possible origin.

VI. Acknowledgments

The author is indebted to Mr. L. L. McCready, Mr. J. D. Murray, and Mr. W. C. Rowe who took part in the observational phase of this work. The author wishes to thank Dr. J. L. Pawsey for helpful criticism and advice in the preparation of this paper.

VII. References

Allen, C. W. (1947).—*Mon. Not. R. Astr. Soc.* **107** : 386-96.

Appleton, E. V., and Hey, J. S. (1946).—*Phil. Mag.* **37** : 73-84.

Bailey, V. A. (1950).—*Phys. Rev.* **78** : 428-43.

Bohm, D., and Gross, E. P. (1949).—*Phys. Rev.* **75** : 1851, 1864.

Bunemann, O. (1950).—*Nature* **165** : 474-6.

Giovanelli, R. G. (1948).—*Nature* **161** : 133-4.

Haeff, A. V. (1949).—*Phys. Rev.* **75** : 1546-51.

Hey, J. S., Parsons, S. J., and Phillips, J. W. (1948).—*Mon. Not. R. Astr. Soc.* **108** : 354-71.

McCready, L. L., Pawsey, J. L., and Payne-Scott, Ruby (1947).—*Proc. Roy. Soc.* A **190** . 357-75.

Martyn, D. F. (1947).—*Nature* **159** : 26-7.
Newton, H. W. (1949).—*Observatory* **69** : 235.
Pawsey, J. L. (1950).—*Proc. Instn. Elect. Engrs.* **97** : 290-308.
Payne-Scott, Ruby, Yabsley, D. E., and Bolton, J. G. (1947).—*Nature* **160** : 256-7.
Ryle, M. (1948).—*Proc. Roy. Soc.* A **195** : 82-97.
Ryle, M., and Vonberg, D. D. (1948).—*Proc. Roy. Soc.* A **193** : 98-120.
Wild, J. P. (1950*a*).—*Aust. J. Sci. Res.* A **3** : 399. (Part II of this Series.)
Wild, J. P. (1950*b*).—*Aust. J. Sci. Res.* A **3** : 541. (Part III of this Series.)
Wild, J. P., and McCready, L. L. (1950).—*Aust. J. Sci. Res.* A **3** : 387. (Part I of this Series.)

EXPLANATION OF PLATE 1

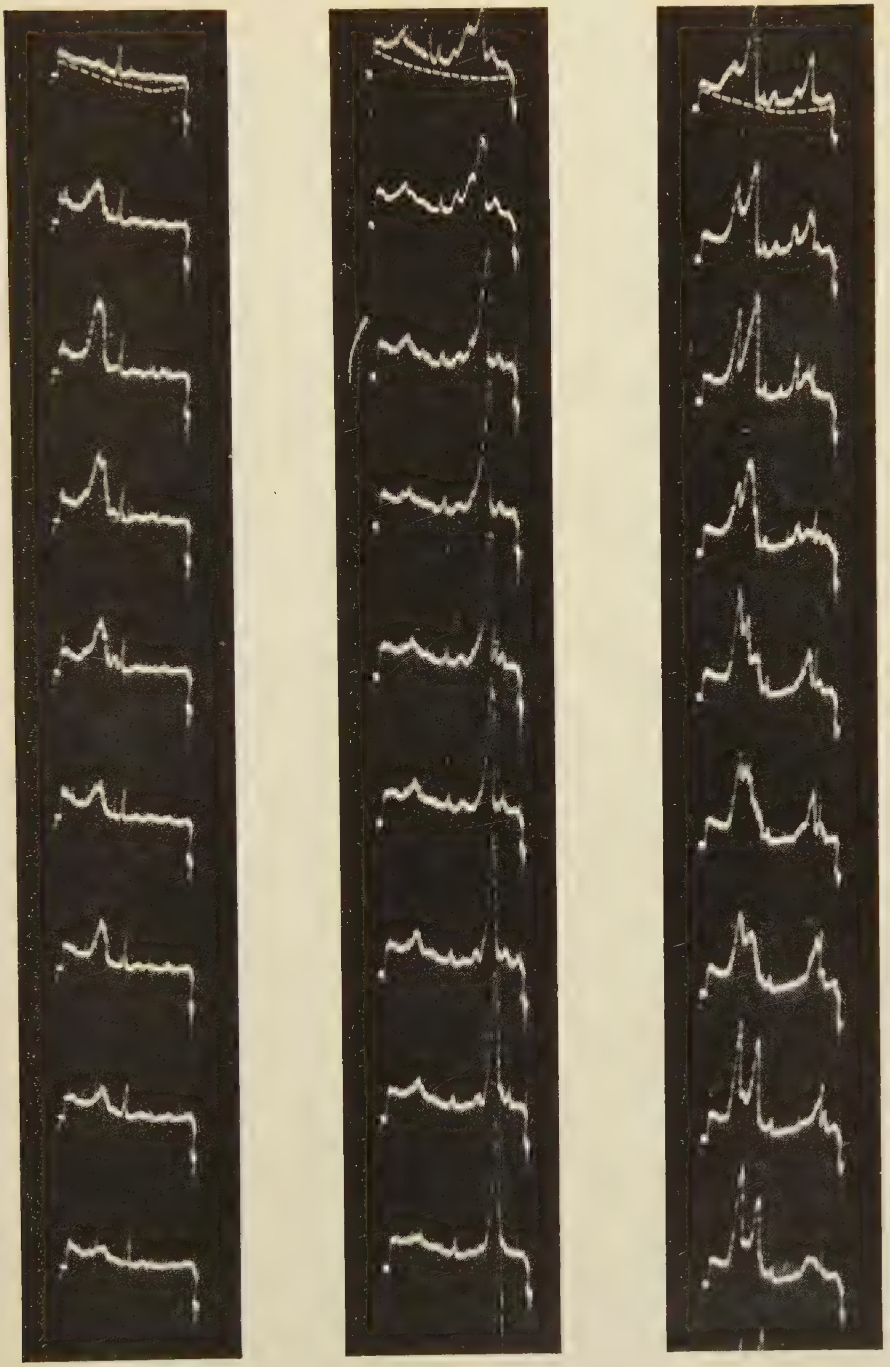
Three sections of record, each of 3 seconds' duration, showing storm bursts. Successive traces are separated by 1/3 second. In each trace, frequency varies continuously in the horizontal direction from about 70 Mc/s. (left-hand edge) to about 130 Mc/s. (right-hand edge), and receiver output in the vertical direction. The dotted lines have been inserted in the first trace of each section to show the level in the absence of solar radiation.

(*a*) A typical storm burst, from start to finish.

(*b*) Part of a large storm burst.

(*c*) Storm bursts during a period of very high activity.

SPECTRUM OF HIGH-INTENSITY SOLAR RADIATION. IV



(a) (b) (c)

RADIO OBSERVATIONS OF TWO LARGE SOLAR DISTURBANCES

By W. N. Christiansen,* J. V. Hindman,* A. G. Little,*
Ruby Payne-Scott,* D. E. Yabsley,* and C. W. Allen†

[*Manuscript received September 29, 1950*]

Summary

On February 17 and 21–22, 1950, two exceptionally large radio-frequency disturbances of solar origin were observed on seven radio receivers working in the frequency range from 62 to 9400 Mc/s. At each of the frequencies the radio-frequency power flux was recorded continuously during the disturbance and at several frequencies the polarization of the radiation was examined. The apparent position of the origin of the radiation was determined at one frequency.

The time of commencement and the duration of the disturbances at the different frequencies were compared with each other and with those of associated phenomena, solar flares, radio fade-outs, and geomagnetic effects. The two disturbances show marked similarities and marked differences ; these are summarized.

I. Introduction

Large outbursts of radio-frequency noise from the Sun, which frequently accompany the more intense solar flares, have been reported in recent years by Appleton and Hey(1946) and others. In the majority of cases the observations were confined to a small number of radio-frequency bands. In order to determine the character of such outbursts more fully, information is required over a wide range of frequency together with information on any other relevant solar or geophysical phenomena.

During February 17 and 21–22, 1950, two very intense outbursts of noise were recorded in Australia on seven frequency bands between 62 and 9400 Mc/s. A wide range of solar and geophysical observations was also available. Taken together these observations give a detailed picture of this type of phenomenon.

Owing to the limited number of outbursts for which detailed observations are available, it is not possible to specify any particular disturbance as being typical. However, a special feature of these two outbursts was the great intensity of the high-frequency radiation. For all frequencies above 600 Mc/s. the power flux was greater than any previously reported.

The observations available covering the time of these outbursts are :

(i) Continuous recording of the power flux on seven radio-frequency bands.

(ii) Some polarization measurements at four of the frequencies.

(iii) Measurements of the apparent position of the source of radio-frequency radiation at one frequency.

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

† Commonwealth Observatory, Canberra, A.C.T.

(iv) Ionospheric fade-out records.
(v) Sunspot sketches.
(vi) Solar flare observations.
(vii) Magnetic records.

II. Observations

(*a*) *Power-Flux Measurements*

The power flux was measured by using calibrated directional aerials and receivers connected to recording milliammeters.

Records on 62, 98, 600, 1200, 3000, and 9400 Mc/s. were obtained at Sydney and the 200 Mc/s. record was taken at Canberra. Cross checks between various equipments lead to an overall probable error of not more than 1·5 db. The recorded values of the power flux per square metre per cycle/sec. bandwidth at each frequency are shown in Figure 1.

(*b*) *Polarization*

The measurements of degree of circular polarization which were made at several frequencies during the disturbances have been summarized in Table 1. The changes indicated in the table were usually not sharp, the transition from

Table 1

SUMMARY OF MEASUREMENTS OF CIRCULAR POLARIZATION—FEBRUARY 17 AND 21–22, 1950

Freq. / U.T.	98 Mc/s.	200 Mc/s.	600 Mc/s.	3000 Mc/s.
0126–0138	Random	—	R.H. up to 50%	—
0138–0146	Random to R.H.* 40%	L.H. 27% at 0144	R.H. up to 50%	—
0146–0200	L.H.* 60%	—	Random to R.H. 30%	—
0200–0500	—	Random at 0505	Random to R.H. 30%	—
–2338	R.H. 100%	—	—	—
2338–2350	Random	—	—	R.H. 13% at 2343
2350–0010	R.H. 100%	—	—	—
0010–0041	L.H. 60%†	Random at 0020 R.H. 4% at 0035	Random	—
0041–0127	Linear**	R.H. 6% at 0114	R.H. varying 0–40%	—
0127–0445	L.H. 100%	R.H. 37% at 0348 R.H. 12% at 0442	R.H. varying 33%	R.H. 17% at 0139

---

* R.H. and L.H. refer to circular polarization unless otherwise stated (radio-electric convention for R. and L.).

† Elliptical polarization.

** Direction of polarization along the axis of spot groups.

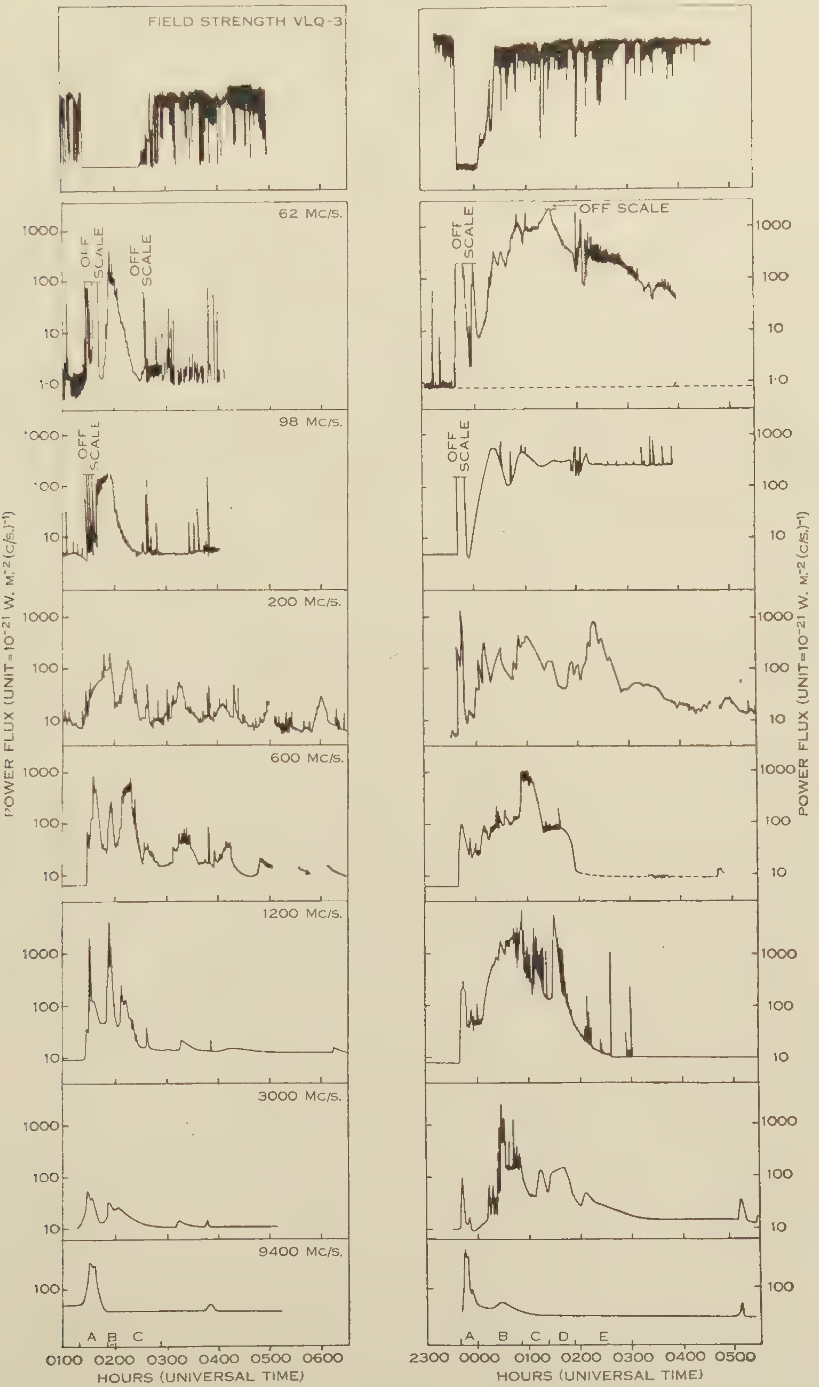


Fig. 1.—Records of power flux for seven frequencies on February 17 and 21–22, 1950. The top record shows field strength of Station VLQ–3, Brisbane, measured at Sydney.

one condition to the next being generally a gradual one. To simplify the presentation of the data continuous information has been omitted, and only the major condition existing over any one period has been shown. The readings at 200 and 3000 Mc/s. were spot readings at times stated.

(*c*) *Position Measurements*

A two-aerial interferometer operating at 97 Mc/s. was used to locate the position of the source of radio-frequency radiation with respect to the solar disk. As these measurements will be described elsewhere by Payne-Scott and Little (in preparation), only the results will be briefly stated here.

Before the commencement of the outburst on February 17 radiation was being received from a source located near the western limb of the Sun. When the outburst commenced at about 0125 hours, a second source appeared at a place on the Sun near which a flare was seen in progress from 0150 hours (see Table 3 below). During the next 30 minutes the position of this source drifted steadily eastwards to a point approximately 10 minutes of arc beyond the Sun's limb.

On February 21–22 the position of the source of radiation was not measured until 0007 hours, some 27 minutes after the start of the outburst. At this time the source was close to the position of the flare which commenced at 2338 hours. During the period from 0040 to 0120 hours the position drifted out to the western limb of the Sun and back to the flare position, where it remained for the next 2 hours, except for three short excursions to the western limb around 0200 hours.

These results are believed to indicate that the origin of the radio-frequency radiation at 97 Mc/s. was in the vicinity of the solar flare on each occasion.

(*d*) *Radio Fade-Outs*

Information concerning the radio fade-outs which accompanied both the disturbances has been obtained from ionospheric recordings, circuit reception data, and field strength recordings at Sydney of Station VLQ–3, Brisbane on 9·6 Mc/s. The field strength recordings give the most accurate timing, and the records have been reproduced at the top of Figure 1.

Fade-out data are summarized in Table 2.

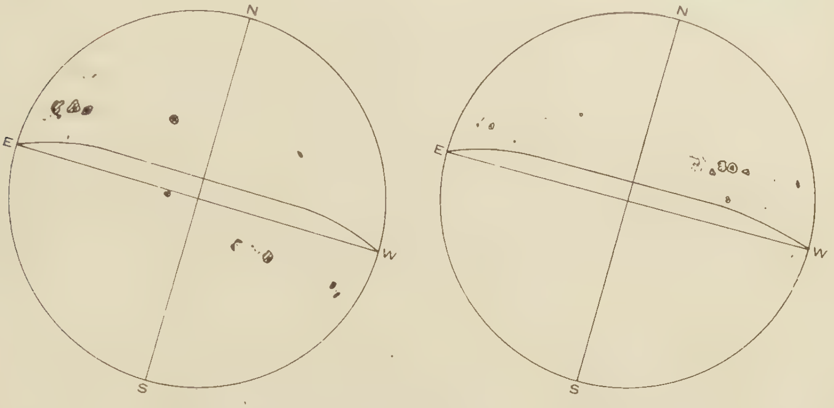
TABLE 2

SUMMARY OF FADE-OUT DATA

Date	Commencement (U.T.)	Finish (U.T.)	Intensity
February 17, 1950	0125 within 1 minute	0240 within 15 minutes	Complete fade
February 21–22, 1950	2341 within 1 minute	0045 within 30 minutes	Complete fade

(*e*) *Sunspots*

Sunspot sketches from the Commonwealth Observatory at Mt. Stromlo (Fig. 2) show that at the time of both outbursts the outstanding feature on the Sun's disk was a complex bipolar group which crossed the central meridian on February 20 with a projected area of 3800-millionths of the visible disk. Several other small groups were also present over this period.



FEBRUARY 16, 1950

FEBRUARY 22, 1950

Fig. 2.—Sunspot sketches for February 16 and 22, 1950, drawn at Commonwealth Observatory, Canberra.

(*f*) *Flares*

Intense flares, which occurred in each case in the vicinity of the large spot group mentioned above, were observed at the time of each of the disturbances. Observations of the flares from Carter Observatory, Wellington, New Zealand, Kodaikanal Observatory, India, and the Commonwealth Observatory, Canberra, are shown in Table 3.

TABLE 3

OBSERVATIONS OF FLARES FROM VARIOUS OBSERVATORIES

Observatory	Date	First Seen (U.T.)	Maximum Intensity (U.T.)	Last Seen (U.T.)	Lat.	Central Meridian Distance	Importance
Carter ..	February 17	0150		0305*	10 °N.	37 °E.	3
Kodaikanal	February 17	2338	0310		—		
Carter ..	February 21–22		2343	0200	13 °N.	33 °W. 24 °W.	3
Canberra	February 22	0000		0040	12 °N.	28 °W.	2
Carter ..	February 22	0257	0302	0331	13 °N.	29 °W.	2

* Times underlined are actual observed times of beginning or end of flare.

The starting and finishing times are in approximate agreement with the fade-outs. It is noted that after 0040 hours, February 22, the Carter Observatory reported considerable flare activity ; this was observed at Canberra, but was regarded as being below Importance 1 threshold.

(*g*) *Magnetic Data*

(i) *Crochet.*—Examination of the magnetic record from Watheroo Magnetic Observatory for February 17, 1950, shows a well-defined crochet starting at about 0124 hours, reaching a maximum at about 0136 hours, and finishing at about 0215 hours U.T.

The Watheroo record is reproduced in Figure 3 and the crochet is represented on Figure 4 for comparison with the other starting times.

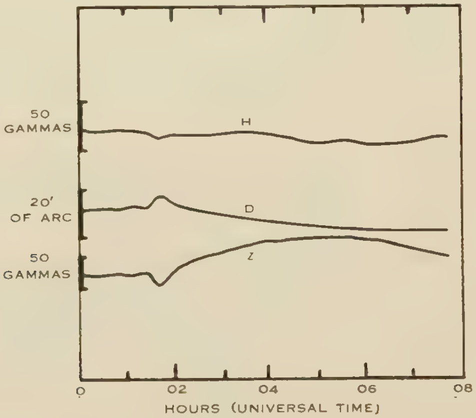


Fig. 3.—Magnetogram for February 17, 1950, taken at Watheroo Magnetic Observatory, Western Australia.

At the maximum of the crochet, the departure of the magnetic field from the pre-crochet value was approximately equal to the departure from the night-time field which was already being produced by the Sun.

On February 21, 1950, the radio-frequency disturbance commenced during a magnetic storm and it is impossible to say with any certainty whether any significant variation occurred on the magnetic record at this time.

(ii) *Storms.*—Two magnetic storms occurred during this period. Table 4 summarizes the data from Watheroo Magnetic Observatory, Western Australia, supplemented by earth-current records of the Commonwealth Observatory.

TABLE 4

Date	Commencement	Type	Max. Phase	Max. *k* Index
February 20, 1950	1700 hours	Gradual	2100 hours	7
February 23, 1950	1041 hours	Sudden	2300 hours	6

It is unlikely that the first storm was associated with the disturbance of February 17, as the delay time of 88 hours is much greater than normal. The gradual commencement indicates that this storm was of the recurrent type. The second storm, on the other hand, was almost certainly associated with the disturbance of February 21–22, the 35-hour delay time being only a little longer than normal.

III. Discussion

(*a*) *Time of Commencement*

Figure 4 shows the starting times of the outbursts plotted for each frequency. The isolated points show the times of preliminary activity, which may or may not be a part of the same disturbance. The main disturbance commenced on each frequency at the point where the heavy line starts.

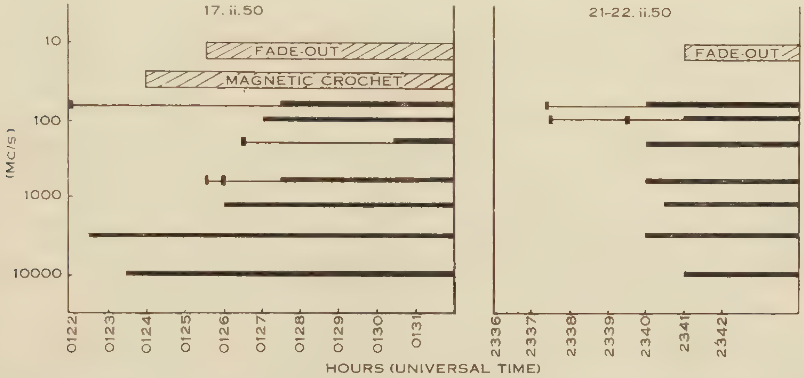


Fig. 4.—Times of start of radio-frequency disturbances on February 17, 1950, and February 21–22, 1950, with start of fade-outs and magnetic crochet.

It is of interest to examine the relative times of the start of the disturbance on different frequencies, as Bolton, Payne-Scott, and Yabsley (1947) and Wild (1950) have shown that there is a tendency for the higher frequencies to start earlier than the lower frequencies, and Covington and Medd (1949) have found that sometimes a high and a low frequency start simultaneously.

Examination of Figure 4 shows that in the disturbance of February 17 the higher frequencies tended to commence earlier. The disturbance of February 21–22 shows very different characteristics, the commencement on all frequencies being almost simultaneous.

(*b*) *Power Flux*

Figure 5 shows the peak intensities reached at each frequency on the two days. Unfortunately the peak intensities reached on 98 Mc/s. and 62 Mc/s. on February 17 and February 21–22 are not known, because the records went off-scale. In spite of this it is quite evident that neither outburst showed a systematic increase or decrease of power with frequency. In both outbursts there was a dip in the peak intensities in the neighbourhood of 200 Mc/s., and a maximum at 1200 Mc/s.

Over a large part of the frequency range each of the bursts under consideration is divided naturally into discrete intervals, during which the flux rises to a maximum and falls again. The disturbance of February 17 is divided into three intervals, *A*, *B*, and *C* in Figure 1, and the average flux over these intervals is given in Table 5. Similar treatment of the disturbance of February 21–22 gives five intervals, *A*, *B*, *C*, *D*, and *E* (Fig. 1), and the average flux for each interval is shown in Table 6. In these tables the maximum average flux on each frequency is shown in heavy type, and it will be seen that there is a tendency for this maximum to occur earlier on the higher frequencies.

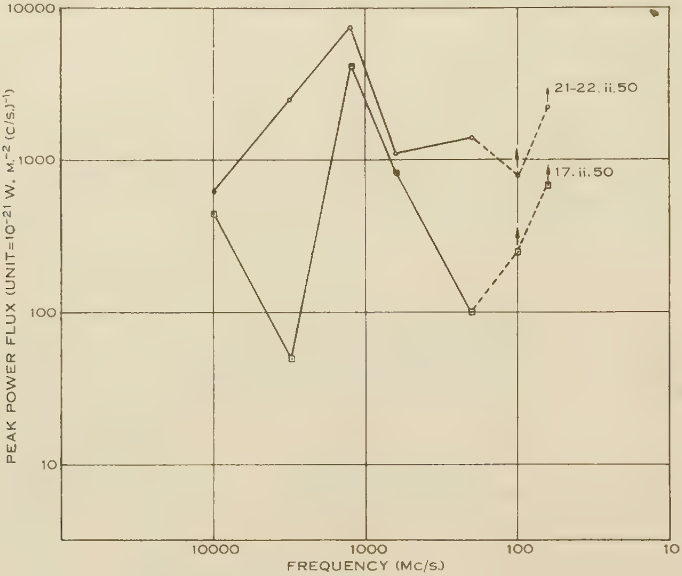


Fig. 5.—Peak power flux *v*. frequency for February 17 and February 21–22, 1950.

TABLE 5.

February 17, 1950	Average Flux (unit$=10^{-21}$W.m.$^{-2}$ (c/s.)$^{-1}$)		
Period / Freq.	*A* 0120–0150	*B* 0150–0202	*C* 0202–0250
9400 Mc/s.	**130**	—	—
3000 Mc/s.	**20**	16	5
1200 Mc/s.	126	**480**	42
600 Mc/s.	122	120	**160**
200 Mc/s.	17	**39**	12
98 Mc/s.	95	**560**	85
62 Mc/s.	>80	**220**	67

TABLE 6

February 21–22, 1950 — Average Flux (unit $=10^{-21}$ W.m.$^{-2}$ (c/s.)$^{-1}$)

Period / Freq.	*A* 2340–0000	*B* 0000–0051	*C* 0051–0121	*D* 0121–0153	*E* 0153–0300
9400 Mc/s.	**230**	13	—	—	—
3000 Mc/s.	25	**293**	77	70	15
1200 Mc/s.	83	**1370**	550	790	15
600 Mc/s.	31	172	**360**	73	10
200 Mc/s.	150	**225**	157	83	220
98 Mc/s.	>80	240	**360**	300	300
62 Mc/s.	>10	205	1050	**1205**	550

It is of interest to note that the 9400 Mc/s. record showed only the first part of the disturbance. In each case a period of several hours elapsed before further bursts were recorded, although the disturbance continued during this interval on all other frequencies—in most cases at a higher intensity.

(*c*) *Energy*

Figure 6 shows the total energy in the disturbance at each frequency integrated over the period of the records ; the peaks and dips of this curve are very similar to those of the peak flux curve. It should be noted that more than two-thirds of the frequency band covered by the observations lies between 3000 and 9400 Mc/s. ; if the energy per cycle bandwidth were the same at all frequencies, most of the energy would lie in this region.

(*d*) *Relationship to Radio Fade-Outs*

At the top of Figure 1 are shown the records of field strength at Sydney from Station VLQ–3, Brisbane, on February 17 and 21–22.

It will be seen from Figure 4 that on February 17 the fade-out commenced after the start of the disturbance on the higher frequencies, and before the start of the disturbances on the frequencies below 1200 Mc/s. On February 17 the duration of the radio-frequency disturbance and the fade-out were roughly equal.

On February 21–22 the fade-out and disturbance occurred practically simultaneously (Fig. 4). On all frequencies on this day the intensity subsided during the period of the fade-out, and then increased to a very high level, in most cases about the time of the end of the fade-out. The correspondence between the end of the fade-out and the increase in intensity is particularly marked on the 3000 Mc/s. record on this day. This suggests that the disturbance on February 21–22 was of a different type from that on February 17, at least from 0000 hours onward. This conclusion is supported by the great variability of the intensity at 3000 Mc/s., in contrast with the usual smooth rise and fall of disturbances at this frequency.

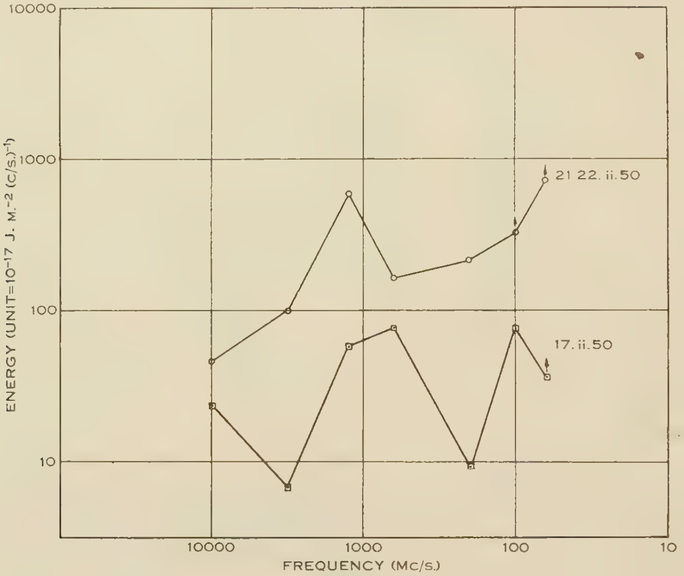


Fig. 6.—Total energy of disturbance *v.* frequency for February 17 and February 21–22, 1950.

IV. Summary of Observations

The general features of the two occurrences are summarized in Table 7. The two disturbances showed marked similarities and marked differences.

Table 7

SUMMARY OF OBSERVATIONS OF THE TWO SOLAR DISTURBANCES

Date Property	February 17, 1950	February 21–22, 1950
Flux	High on all frequencies	High on all frequencies
Commencement	High frequencies first, low frequencies later	Simultaneous all frequencies
Flare	Flare observed	Flare observed
Fade-out	Fade-out during time of disturbance	Fade-out during first part of disturbance
Magnetic storm ..	Doubtful	35 Hours later
Magnetic crochet ..	Present	Doubtful
Circular polarization ..	Low frequency changes of sense High frequency fixed sense	Low frequency changes of sense High frequency fixed sense
Power distribution ..	Maximum power earlier at high frequencies	Maximum power earlier at high frequencies
Duration	Similar at all frequencies	Longer at low frequencies

Whether there are distinctive types of disturbances to which one or both of these may belong is a question which can only be answered after analysis of large numbers of disturbances—perhaps along the lines indicated in Table 7.

V. Acknowledgments

Work of this nature necessarily involves a large number of people in observation, and we have to acknowledge our indebtedness particularly to Mr. G. Fairweather, Mr. C. F. Fryar, Mr. A. J. Harragon, and Mr. G. H. Trensky of the Radiophysics Laboratory, and Miss Joyce Daly of the Commonwealth Observatory. We wish to record our appreciation of the assistance of Dr. J. L. Pawsey, Assistant Chief, Division of Radiophysics, C.S.I.R.O. We are indebted to Mr. J. S. Rayner, Chief Geophysicist, Bureau of Mineral Resources, Geology and Geophysics, and to Mr. L. S. Prior for copies of Magnetograms from the Bureau's observatories at Toolangi and Watheroo. Solar flare information was kindly supplied by Dr. J. L. Thomsen, Director of the Carter Observatory, Wellington, New Zealand, and by Dr. A. K. Das, Director of the Solar Physics Observatory, Kodaikanal, India.

VI. References

Appleton, E. V., and Hey, J. S. (1946).—*Phil. Mag.* **37** : 73-84.
Bolton, J., Payne-Scott, Ruby, and Yabsley, D. E. (1947).—*Nature* **160** : 256.
Covington, A. E., and Medd, W. J. (1949).—*J.R. Astr. Soc. Can.* **43** : 106-10.
Wild, J. P. (1950).—*Aust. J. Sci. Res.* A **3** : 541-57.

THERMOELECTRIC PROPERTIES OF TITANIUM WITH SPECIAL REFERENCE TO THE ALLOTROPIC TRANSFORMATION

By H. W. Worner*

[*Manuscript received September* 26, 1950]

Summary

A differential method for determining the thermoelectric power of titanium/platinum thermocouples is described. The results of the differential measurements are compared with data obtained in direct e.m.f. *versus* temperature determinations. From the thermoelectric power results, the temperature dependency of the following properties has been deduced: (*a*) the Peltier coefficient of the titanium/platinum couple; (*b*) the difference between the Thomson coefficients of titanium and platinum; (*c*) the absolute thermoelectric power of titanium. It is shown that abrupt changes in thermoelectric properties accompany the α-β transformation in refined titanium, the transition temperature range being 883 ± 1 to 887 ± 2 °C. In a commercial grade of titanium, the transformation occurs over the range 865 to 970 °C. Some effects due to preferred orientation and the elimination thereof are also reported.

I. Introduction

The investigation described herein was undertaken primarily to throw further light on the allotropic transformation in titanium. Studies of other properties including electrical resistance (de Boer, Burgers, and Fast 1936), magnetic susceptibility (Squire and Kaufmann 1941), and solubility of hydrogen (McQuillan 1950) have yielded information concerning the transformation temperature range in samples of different degrees of purity. However, it was considered that an investigation of the allotropic change by means of another physical property such as thermoelectric power would be of value in determining the transformation temperature and the effect of impurities thereon. It was also thought that thermoelectric power determinations might be of value in ascertaining $\alpha+\beta$ ranges in titanium-base alloys.

Greiner and Ellis (1949) have reported on the thermoelectric properties of couples comprising platinum and titanium of "technical" purity, that is, titanium made by reducing titanium tetrachloride with magnesium. These workers covered the range -200 to 1000 °C., but they apparently did not attempt to reveal any effects related to the allotropic transformation. Also, the results of Greiner and Ellis did not provide any definite evidence concerning the possible effects of preferred orientation in titanium wire. In view of the anisotropy of electrical properties in hexagonal metals, such effects might well be expected in the case of titanium and they were therefore sought in the course of the present investigation.

---

* Physical Metallurgy Section, C.S.I.R.O., Baillieu Laboratory, University of Melbourne.

II. Experimental

(*a*) *Metals Used*

Titanium refined by the iodide dissociation method in the Philips Laboratories, Eindhoven, was employed in most of the experiments. The content of impurities in this metal are indicated in Table 1. The hardness of the refined titanium after annealing for 1 hour at 850 °C. was 70 Vickers units. This low hardness was viewed as confirmation of the analytical results in respect to oxygen and nitrogen, both of which are known to harden titanium appreciably even in concentrations of the order of 0·1 at.–%.

Table 1

CONTENT OF IMPURITIES IN REFINED AND COMMERCIAL-GRADE TITANIUM

(Concentrations in atomic per cent.)

Element	Method of Analysis	Refined Titanium	Commercial Grade Titanium
Carbon	Chemical	—	0·3
Nitrogen	Chemical	0·03	0·4
Oxygen	Chemical	0·02	0·9
Copper	Chemical	—	<0·02
	Spectrographic	Faint trace	—
Magnesium	Spectrographic	Very faint trace	Faint trace
Zinc	Spectrographic	Not detected	Faint trace
Silicon	Chemical	—	<0·04
	Spectrographic	Faint trace	Trace
Tin	Spectrographic	Strong trace	Trace
Lead	Spectrographic	Not detected	Trace
Vanadium	Spectrographic	Faint trace	Not detected
Antimony	Spectrographic	Strong trace	Not detected
Manganese	Spectrographic	Very faint trace	Trace
Iron	Chemical	0·02	0·3
Cobalt	Spectrographic	Not detected	Faint trace
Na, K, Ag, Be, Ca, Sr, Ba, Cd, B, Al, Zr, Nb, Ta, As, Bi, Cr, Mo, W, Ni	Spectrographic	Not detected	Not detected

Portions of the bar of refined metal were rolled into strips ranging in thickness from 0·2 mm. down to 0·08 mm. The metal required only two intermediate annealing treatments of about ½ minute at 550 to 600 °C. during the reduction of the ¼ inch diameter bar down to the final thicknesses. After the second anneal, the thin oxide film was removed from the surface by abrasion with fine carborundum paper. By means of metal shears, long ribbons of titanium, about 0·5 to 1 mm. wide, were prepared for convenience in making thermocouple assemblies. Prior to use, the ribbons were cleansed by rubbing with fine, waterproof carborundum paper. Finally, they were wiped with a cloth soaked in carbon tetrachloride to remove grease films resulting from handling. Thereafter, every effort was made to prevent further contamination of the surface of the strip.

Some commercial-purity titanium made by the Kroll process (Worner 1949) was produced in the form of 0·5 mm. diameter wire and the latter used to obtain results for comparison with those pertaining to refined metal. The contents of impurities in the commercial grade of metal are listed in Table 1 alongside those for the refined titanium. The hardness of the commercial metal after annealing for $1\frac{1}{2}$ hours at 850 °C. was 230 Vickers units (cf. 70 for refined metal).

Platinum of thermocouple purity was used as the reference metal for studying the thermoelectric properties of titanium. Temperatures and temperature differentials were measured by couples comprising platinum and 13 per cent. rhodium-platinum alloy, both of thermocouple purity. The noble metal wires were 0·25 mm. in diameter. Prior to use, they were heated to about 1400 °C. for $\frac{1}{2}$ hour by passage of an alternating current.

(b) *Outline of Methods of Measurement*

The research involved two main groups of measurements :

(i) Direct determinations of e.m.f. *versus* temperature relationships for titanium/platinum couples.

(ii) Determination of the variation of thermoelectric power of titanium/platinum couples with temperature, the measurements being made by a differential method which was developed during the course of the research.

The results of the differential method (ii) could be employed to calculate e.m.f.–temperature relationships since

$$\varphi \left| \begin{matrix} T \\ T_0 \end{matrix} \right. = \int_{T_0}^{T} P.dT,$$

where $\varphi \left| \begin{matrix} T \\ T_0 \end{matrix} \right.$ is the e.m.f. developed in a titanium/platinum thermocouple circuit with junctions at T and T_0 and P is the thermoelectric power of the couple. The calculated values of φ (which were actually obtained by graphical integration) could be compared with the experimentally determined values so that the two groups of measurements, (i) and (ii) above, could be checked against one another.

(c) *Direct e.m.f.* versus *Temperature Determinations*

A triple thermocouple comprising titanium, platinum, and 13 per cent. rhodium-platinum alloy was employed in these experiments. At the hot junction, the three components were brought together so that they were in thermal as well as electrical contact. This was done so that there would be little risk of a difference in temperature between the titanium · platinum junction and the reference junction of platinum and rhodium-platinum. In this connection, it should be noted that the hot junction was located at the centre of a 7 cm. span of constant temperature in a long, 2·5 cm. diameter tube furnace.

The general assembly of the junctions, furnace, vacuum envelope, etc. is shown diagrammatically in Figure 1. At the hot junction (*A*), direct contact between the titanium and the rhodium-platinum · platinum junction could not be tolerated, otherwise the latter would become contaminated with titanium by diffusion. Therefore, prior to each series of measurements, the rhodium-platinum · platinum junction was wrapped in two layers of 0·13 mm. thick platinum sheet and the end of the titanium lead was bound on to the platinum wrapping by two or three turns of fine platinum wire.

In the hot zone, the leads were held in aluminous porcelain insulators, while outside the furnace, silica insulators were used. Chief reason for using the porcelain in the hot zone was that titanium suffered less superficial contamination when in contact with aluminous porcelain than when alternative materials like silica and alumina were used. The leads emerged from the vacuum envelope through porcelain insulators set in a brass flange at the open end of the silica vacuum envelope. To provide vacuum seals, Apiezon wax was melted into the spaces between the wires, porcelain insulators, and the flange. Surrounding the hot junctions and the thermocouple leads within the furnace, there was an earthing shield consisting of thin molybdenum sheet in the hot zone where temperatures were likely to be higher than 800 °C., and nickel foil (0·05 mm. thick) in the zone of steep temperature gradient.

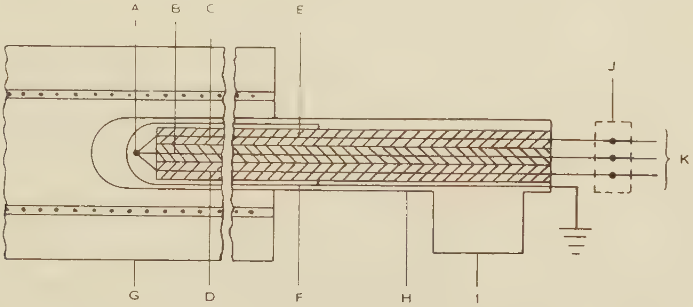


Fig. 1.—Diagrammatic representation of equipment used in e.m.f. *versus* temperature determinations.
A, Hot junction ; *B*, titanium lead ; *C*, platinum lead ; *D*, rhodium-platinum lead ; *E*, insulators ; *F*, earthing shield ; *G*, nichrome resistance furnace ; *H*, fused silica vacuum envelope ; *I*, connection to vacuum pump ; *J*, cold junctions (melting ice) ; *K*, leads to potentiometer.

To minimize superficial contamination of the titanium with oxygen, nitrogen, and carbon during experiments at temperatures above 300 °C., the pressure in the vacuum envelope had to be kept below 10^{-4} mm. mercury. This was achieved by means of an Apiezon oil diffusion pump and an elongated, wide-bore coupling which served as a vapour trap. Pressures were measured by a Philips ionization gauge.

The cold junctions between the thermocouple components and fine copper potentiometer leads were accommodated in thin-walled glass sheaths filled with pure liquid paraffin. The oil prevented condensation of water on the cold wires and so precluded the possibility of electrochemical effects causing errors. A mixture of crushed ice and water was maintained around the cold junction sheaths and the cold junctions themselves were about 18 cm. below the surface of the ice-water mixture.

Measurements of e.m.f. were made only when the hot junction temperature was practically constant. One potentiometer was used ; hence the two e.m.f.'s required at each temperature could not be measured simultaneously. For this reason, the following sequence of measurements was made at any given temperature :

(i) e.m.f. of rhodium-platinum · platinum couple.
(ii) e.m.f. of titanium · platinum couple.
(iii) Repeat (i).

The results were accepted only when thermal conditions were so steady that (iii) equalled (i) within the limits of accuracy attainable.

A Cambridge 0 to 100 millivolt range potentiometer was used in the great majority of the observations. Some careful check measurements were made from time to time by means of a Leeds and Northrup, White double-bridge potentiometer which could be used to measure e.m.f.'s to the nearest microvolt.

(*d*) *Determination of Thermoelectric Power by Differential Method*

A differential method for determining thermoelectric power at temperatures up to 1100 °C. was developed because it proved impossible to deduce, with sufficient accuracy, values of thermoelectric power from the e.m.f.–temperature results. A system of differential thermocouples which was comparatively simple to construct and which proved very convenient in use is illustrated diagrammatically in the upper part of Figure 2. In essence it consisted of a platinum/titanium/platinum differential couple ($\mathrm{Pt}_{(a)} \cdot \mathrm{Ti} \cdot \mathrm{Pt}_{(b)}$ in Fig. 2), and a platinum/rhodium-platinum/platinum differential couple ($\mathrm{Pt}_{(c)} \cdot \mathrm{Rh\,Pt}_{(d)} \cdot \mathrm{Pt}_{(b)}$ in Fig. 2) having one common platinum lead $\mathrm{Pt}_{(b)}$) and consequently a triple junction ($\mathrm{Ti} \cdot \mathrm{Pt}_{(b)} \cdot \mathrm{Rh\,Pt}_{(d)}$). Provided both the junctions $\mathrm{Ti} \cdot \mathrm{Pt}_{(a)}$ and $\mathrm{Rh\,Pt}_{(d)} \cdot \mathrm{Pt}_{(c)}$ were at exactly the same temperature, T_2, while the triple junction was at T_1 (T_1 being about 2 to 5 °C. less than T_2), then

$$\frac{\triangle\varphi_{\mathrm{Ti/Pt}}}{\triangle\varphi_{\mathrm{Rh\,Pt/Pt}}}=\frac{P_{\mathrm{Ti/Pt}}}{P_{\mathrm{Rh\,Pt/Pt}}},$$

where $\triangle\varphi_{\mathrm{Ti/Pt}}$=e.m.f. of $\mathrm{Pt}_{(a)}/\mathrm{Ti}/\mathrm{Pt}_{(b)}$ differential couple.
$\triangle\varphi_{\mathrm{Rh\,Pt/Pt}}$=e.m.f. of $\mathrm{Pt}_{(c)}/\mathrm{Rh\,Pt}_{(d)}/\mathrm{Pt}_{(b)}$ differential couple, and
$P_{\mathrm{Ti/Pt}}$=thermoelectric power of the titanium/platinum couple at $(T_1+T_2)/2$,

and similarly for $P_{\mathrm{Rh\,Pt/Pt}}$. Values for the latter were available in standard tables. Hence to get $P_{\mathrm{Ti/Pt}}$, it was simply necessary to determine $(\triangle\varphi_{\mathrm{Ti/Pt}})/(\triangle\varphi_{\mathrm{Rh\,Pt/Pt}})$ which from an electrical standpoint proved easier than actually determining the differential e.m.f.'s separately. The circuit of the special potentiometer developed for determining the ratios of the small e.m.f.'s is outlined in Figure 2 (Items *C* to *I* inclusive). Details concerning the construction and use of this potentiometer will be given in the next section.

The temperatures T_1 and T_2 at the extremities of the differential couples were ascertained to ±0·2 °C. by using a 0 to 100 millivolt range potentiometer (J in Fig. 2) to measure the e.m.f.'s of the couples $\mathrm{Rh\,Pt}_{(d)\ \&\ (e)}/\mathrm{Pt}_{(b)}$ and $\mathrm{Rh\,Pt}_{(e)}/\mathrm{Pt}_{(c)}$.

It was desirable to ensure that junctions $\mathrm{Ti}\cdot\mathrm{Pt}_{(a)}$ and $\mathrm{Rh\,Pt}_{(e)}\cdot\mathrm{Pt}_{(c)}$ did not differ in temperature by more than 0·02 °C. With this aim in view, the following provisions were made :

(i) The differential couples were located at the centre of a long tube furnace where the junctions $\mathrm{Ti}\cdot\mathrm{Pt}_{(a)}$ and $\mathrm{Rh\,Pt}_{(e)}\cdot\mathrm{Pt}_{(c)}$ were near the mid-point of a 7 cm. long zone in which the temperature was constant within ±0·25 °C. This minimized heat loss from the junctions by conduction along the thermocouple wires themselves.

(ii) The two junctions lay in a plane perpendicular to the axis of the tube furnace, the distance between the junctions being only 1·5 mm.

(iii) Surrounding the two junctions and the thin-walled porcelain insulators which supported them, there was a hollow cylinder of tantalum, 3·5 cm. long by 6 mm. diameter by 0·2 mm. wall thickness. The junctions lay close to the mid-point of this tantalum cylinder, where black-body radiation conditions were approximated.

The distance between the triple junction and the other two junctions was 8 cm. The triple junction was located in a zone of shallow temperature gradient in the tube furnace. To minimize this gradient in the immediate vicinity of the junction a tantalum cylinder, similar to that described in the previous paragraph, was placed around the triple junction. However, there was scarcely any need to take such special precautions because the three leads were in thermal, as well as electrical, contact. Since tantalum acts as a " getter " at temperatures above 800 °C., the tantalum cylinders served to aid in the prevention of superficial contamination of the titanium during high temperature measurements.

The triple junction was made in the manner described in the section dealing with direct e.m.f.–temperature determinations. This method of constructing the junction facilitated the connection and removal of titanium specimens as well as preventing contamination of the $\mathrm{Rh\,Pt}_{(d)}\cdot\mathrm{Pt}_{(b)}$ junction. The junction $\mathrm{Ti}\cdot\mathrm{Pt}_{(a)}$ was made by overlapping the ends of the platinum and titanium to the extent of about 1 mm. and binding the overlap with a couple of turns of platinum wire. Whenever a specimen was removed, that portion of the platinum which had been in contact with the titanium was cut off so that uncontaminated platinum was available for connection to the next specimen.

The other details of the thermocouple assembly were designed so that a specimen could be removed and replaced without any derangement of the platinum and rhodium-platinum leads and their junctions. During the initial connection of the leads $\mathrm{Pt}_{(b)}$, $\mathrm{Rh\,Pt}_{(d)\ \&\ (e)}$, and $\mathrm{Pt}_{(c)}$ (all of which had been annealed at 1400 °C. for $\frac{1}{2}$ hour), the wires were kept straight except in the immediate vicinity of the junctions. Before connecting the first titanium specimen, the junctions $\mathrm{Rh\,Pt}_{(d)\ \&\ (e)}\cdot\mathrm{Pt}_{(c)}$ and $\mathrm{Rh\,Pt}_{(d)}\cdot\mathrm{Pt}_{(b)}$ were annealed by means of an oxygen-hydrogen flame. Thereafter care was taken not to distort any of the noble metal leads or their junctions.

It was necessary to accommodate the whole differential thermocouple assembly in a long, silica vacuum envelope in which the pressure could be kept below 10^{-4} mm. mercury. Also, it was essential to provide an earthing shield around the thermocouple components within the furnace. The constructional details necessitated by these requirements were similar to those already described in the section concerned with direct e.m.f.–temperature measurements. The cold junction arrangements described in this section were also used in the differential experiments.

(*e*) *Ratio Potentiometer*

The potentiometer for measuring the ratio $(\triangle\varphi_{\mathrm{Ti/Pt}})/(\triangle\varphi_{\mathrm{Rh\,Pt/Pt}})$ was built up especially for this research because it was not possible to modify any standard potentiometer for the purpose. The essential details of the potentiometer are set out in Figure 2. In brief, the main stages in the use of the potentiometer were as follows. By setting the tapping switch *E* in position 3, $\triangle\varphi_{\mathrm{Rh\,Pt/Pt}}$ was applied across the 50·0 ohm resistance in the set of balancing resistances *F*. The potentiometer current flowing from the 2-volt accumulator *I* was adjusted by means of the high-resistance rheostat *G* so that the potential drop across the 50·0 ohm resistance exactly balanced $\triangle\varphi_{\mathrm{Rh\,Pt/Pt}}$. Then, with switch *E* in

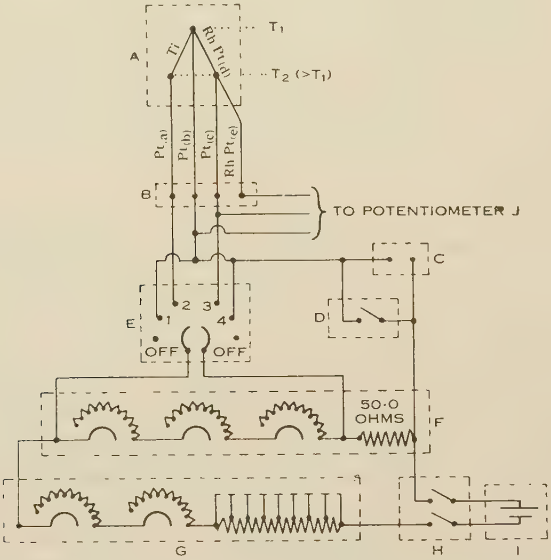


Fig. 2.—Circuit diagram of differential thermocouple assembly and ratio potentiometer.

A, Nichrome resistance tube furnace ; *B*, cold junction box (melting ice) ; *C*, high sensitivity galvanometer ; *D*, switch for protecting *C* ; *E*, tapping switch ; *F*, potentiometer balancing resistances—including one 50·0 ohm resistance and three decade resistances to give 0 to 55·5 ohms in steps of 0·05 ohm; *G*, rheostat—to provide up to 8×10^{6} ohms in steps of $2\cdot5\times10^{3}$ ohms ; *H*, accumulator switch ; *I*, 2-volt lead accumulator ; *J*, potentiometer for measuring temperatures T_1 and T_2.

position 2, and still the same setting of rheostat G, the variable resistances in the balancing series F were adjusted to give a potential drop equal to $\triangle\varphi_{\mathrm{Ti/Pt}}$. Since the variable resistances in F were exceedingly small compared with the resistances required in rheostat G, variations in F for the second potentiometric balance would cause a negligible change in the potentiometric current. Hence, if the balancing resistance for the balance against $\triangle\varphi_{\mathrm{Ti/Pt}}$ were $(50\cdot0+x)$ ohms, then

$$\frac{\triangle\varphi_{\mathrm{Ti/Pt}}}{\triangle\varphi_{\mathrm{Rh\,Pt/Pt}}}=\frac{50\cdot0+x}{50\cdot0}.$$

The value of x could be varied by means of the three decade tapping switches shown diagrammatically in F (Fig. 2) so that the ratio could be read directly off the dials.

It should be appreciated that the actual values of $\triangle\varphi$ were only of the order 20 to 80 microvolts so that certain electrical precautions had to be taken, especially in regard to (*a*) the prevention of stray currents in the potentiometer circuit and (*b*) allowing for parasitic e.m.f.'s.

The accumulator I, switch H, and rheostat G were mounted on polystyrene to prevent any significant leakage of current. In some early experiments, bakelite was used in mounting the switch H and the rheostat G, but for any particular setting of G, the potentiometer current was found to vary appreciably from minute to minute, particularly during humid weather. The use of polystyrene reduced this problem to negligible proportions. It also proved necessary to exercise special care in insulating the leads connecting all parts of the circuit. The resistances in G were all of the high-stability, wire-wound type and they were accommodated in a wooden box to prevent rapid temperature changes.

The balancing resistances F were a set of Wolff Kelvin double bridge resistances. The tapping switches in F were of massive construction in brass and the sudden changes in parasitic e.m.f.'s generated by movement from one setting to another were only of momentary duration. Also, provided the temperature of the whole unit F was kept constant, the total parasitic e.m.f. measured across the whole series of resistances for any particular setting of the tapping switches was not significantly different from that found at any other setting. This constancy of parasitic e.m.f.'s in the balancing resistances was an important factor contributing to the success of the method from an electrical standpoint. Of course, all parts of F, especially the external contacts, had to be thermally insulated very carefully so that there would be no appreciable temperature change during the time required for establishing the balances involved in the measurements.

In view of the unavoidable existence of small parasitic e.m.f.'s in the series of balancing resistances F, the galvanometer connections, and in the tapping switch E, the latter was provided with two positions, 1 and 4, by means of which the galvanometer deflections could be observed when only the parasitic e.m.f.'s were acting. Naturally these observations were made when switch H was open. The resting point of the galvanometer with switch E at position 4 was used as the "zero" when reaching the balance against $\triangle\varphi_{\mathrm{Rh\,Pt/Pt}}$. Likewise the galvanometer resting point with E at position 1 was used in achieving the balance against

$\triangle\varphi_{Ti/Pt}$. By these means, errors due to parasitic e.m.f.'s were practically eliminated provided, of course, that in the switch E, the parasitic e.m.f. in position 1 equalled that in position 2 and similarly for positions 3 and 4. These latter requirements were achieved by (*a*) the use of a rotating brush-type switch with massive contact segments all made from the same billet of brass, and (*b*) adequate thermal insulation to prevent temperature differences within the whole switch unit E. The temporary changes in parasitic e.m.f. due to movement of the contact brushes from one position to another lasted only 1 or 2 seconds. Here it might be noted that with a view to protecting the high-sensitivity galvanometer C, switch D was closed during alterations in the tapping positions in E.

(*f*) *Checking the Differential Method*

It has been indicated that efforts were made to eliminate sources of inaccuracy in the differential method. Nevertheless, the equipment and the procedure for using the potentiometer were subjected to check tests which were chiefly concerned with :

(i) Constancy of parasitic e.m.f.'s in the circuit comprising the balancing resistances F, tapping switch E, and galvanometer connections.

(ii) Effectiveness of insulation throughout the whole circuit.

(iii) Equality of temperature at the junctions $\mathrm{Ti}\cdot\mathrm{Pt}_{(a)}$ and $\mathrm{Rh\,Pt}_{(d)\,\&\,(e)}\cdot\mathrm{Pt}_{(c)}$.

The testing involved in (i) and (ii) above was comparatively simple and it was carried out regularly during the course of the research. In connection with (iii), it might seem that an auxiliary rhodium-platinum lead should have been connected to junction $\mathrm{Ti}\cdot\mathrm{Pt}_{(a)}$ so that the difference between the two temperatures could be ascertained. Actually this was done in a preliminary experiment during the development of the differential method and it was found that the temperature difference was scarcely detectable, even when a high sensitivity galvanometer was employed to reveal the e.m.f. when the two rhodium-platinum/platinum couples were connected in opposition to one another. Thereafter, the auxiliary rhodium-platinum lead was eliminated to reduce the lateral dimensions of the differential thermocouple assembly.

An indirect method of revealing lack of equality between the temperatures of junctions $\mathrm{Rh\,Pt}_{(d)\,\&\,(e)}\cdot\mathrm{Pt}_{(c)}$ and $\mathrm{Ti}\cdot\mathrm{Pt}_{(a)}$ was to ascertain whether the value of $(\triangle\varphi_{Ti/Pt})/(\triangle\varphi_{Rh\,Pt/Pt})$ varied when the temperature differential (T_2-T_1) was varied. For any given value of the mean temperature $(T_1+T_2)/2$, the ratio of e.m.f.'s should remain constant when T_2-T_1 is varied within a range of a few degrees C. It is unlikely that this would hold if the value of T_2 at $\mathrm{Ti}\cdot\mathrm{Pt}_{(a)}$ were significantly different from that at $\mathrm{Rh\,Pt}_{(d)\,\&\,(e)}\cdot\mathrm{Pt}_{(c)}$. Experiments aimed at testing the constancy of $(\triangle\varphi_{Ti/Pt})/(\triangle\varphi_{Rh\,Pt/Pt})$ for various values of (T_2-T_1) were conducted from time to time and a typical set of results is presented graphically in Figure 3. In the range $(T_2-T_1)=0\cdot9$ to $1\cdot9$ °C., the values were scattered and they tended to be higher than those between 2 and $7\cdot3$ °C. The scatter in the results for low values of (T_2-T_1) was largely attributable to lack of sensitivity in establishing the potentiometer balances. The higher values of the ratio in the range $T_2-T_1=0\cdot9$ to $1\cdot9$ °C. were probably due to a slight but significant difference between the temperatures of the $\mathrm{Ti}\cdot\mathrm{Pt}_{(a)}$ and $\mathrm{Rh\,Pt}_{(d)\,\&\,(e)}\cdot\mathrm{Pt}_{(c)}$ junctions. All the results recorded later in this paper were obtained when (T_2-T_1) lay between $2\cdot5$ and 6 °C., but mostly a temperature differential of about 3 to 4 °C. was employed.

The above experiments revealed another important aspect of the method, namely, the importance of allowing sufficient time for steady thermal conditions to be achieved before making measurements. The complete sequence of operations involved in finding the galvanometer resting points, establishing the two potentiometer balances and measuring the temperatures T_2 and T_1 required approximately 2 minutes. It was found that the potentiometer balances were difficult to establish accurately unless the rate of change of T_2 and T_1 was less than $0\cdot1$ °C. per minute. For this reason it proved best to provide the resistance furnace with power from an auto-transformer by means of which any required current could be maintained for as long as desired. During check experiments, it was found that the most convenient way of controlling the temperature differential (T_2-T_1) was to adjust the position of a thin-walled, hollow cylinder of nickel which was placed inside the tube furnace so as to withdraw heat from the vicinity of the triple junction.

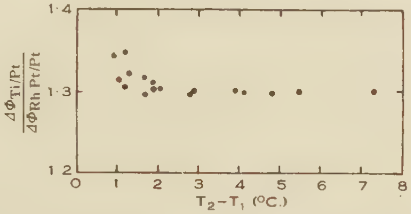


Fig. 3.—Values of $(\triangle\varphi_{\mathrm{Ti/Pt}})/(\triangle\varphi_{\mathrm{Rh\,Pt/Pt}})$ for $(T_1+T_2)/2=$ 805 to 806 °C., plotted against (T_2-T_1).

While the differential method was experimentally analysed and checked in the ways outlined above, there was yet another, and really more comprehensive, way in which the method was tested. As indicated in an earlier section, the results of the differential method could be employed to calculate the e.m.f. *versus* temperature relationship for the titanium/platinum couple. Comparison of the calculated relationship with that determined by direct experiments would then serve as a check on the differential method. This has been done and the observations are reported in Section III.

(*g*) *Calibration of Rhodium-Platinum/Platinum Thermocouples*

Samples of rhodium-platinum/platinum thermocouple wires from each of two batches employed in the course of this investigation were calibrated at the steam point and the lead, aluminium, and silver freezing points. The silver was of $99\cdot98$ per cent. purity and was contained in a graphite crucible furnished with a graphite lid to ensure adequately reducing conditions during melting and solidification. The aluminium was of $99\cdot98$ per cent. purity and the lead of 99·99 per cent. purity. These metals were also melted in a covered graphite crucible. During the calibrations, the thermocouple wires were supported in aluminous porcelain insulators. In the freezing-point calibrations, direct e.m.f. *versus* time measurements were made, the cooling rate being approximately 2·5 °C. per minute prior to the commencement of solidification.

The e.m.f.'s measured at the various standard points are listed in Table 2. At the lead, aluminium, and silver points, the e.m.f.'s of the couples were approximately 0·1 per cent. lower than those indicated in the reference table (Barber 1950) used in the investigation. For all temperatures in the ranges between the metal freezing points, it was assumed that the e.m.f.'s of the couples used were 0·1 per cent. lower than those indicated in the reference table.

Table 2

CALIBRATION OF RHODIUM-PLATINUM/PLATINUM THERMOCOUPLES
Cold junction temperature = 0 °C.
Temperatures in °C., e.m.f.'s in millivolts

Standard Point	e.m.f. from Reference Table (Barber 1950)	Calibration Results	
		Batch I	Batch II
Steam point (758 mm.; 99·9 °C.) ..	0·6433	0·643	0·643
Lead (327·4)	2·660	2·657	2·656
Aluminium (660·0)	6·263	6·257	6·260
Silver (960·8)	9·993	9·978	9·980

During the course of the thermoelectric power determinations, it was noted that slight vaporization of titanium occurred above 900 °C. (cf. Section III (*f*)). There was therefore a risk of the platinum and rhodium-platinum leads and junctions becoming contaminated with titanium. However, recalibration of a used rhodium-platinum couple indicated that the effects of any contamination which may have occurred were not detectable. Apparently the measures adopted to prevent contamination at and near the junctions by direct contact served also to prevent any significant contamination by titanium vapour.

III. Results

(*a*) *Condition of Specimens*

Apart from certain results to be presented towards the end of this portion of the paper, the results pertain to vacuum-annealed, polycrystalline titanium in which there was little or no preferred orientation. X-ray diffraction experiments showed that there was some degree of preferred orientation in the strip produced by rolling the refined titanium bar. The basal planes tended to lie parallel to the rolling plane. However, it was found that there was none but the slightest residual preferred orientation in the ribbon after heating above the α–β transformation range. Therefore it was made standard practice to heat the strip in the interval 900 to 950 °C. for at least $\frac{1}{4}$ hour prior to the commencement of measurements. Apart from eliminating preferred orientation, the heating into the β region almost certainly ensured complete elimination of other effects due to cold working.

(*b*) *Direct e.m.f.-Temperature Determinations*

Before the results are considered, it should be noted that the titanium lead was made by spot-welding together some straight, 22-cm. lengths of refined

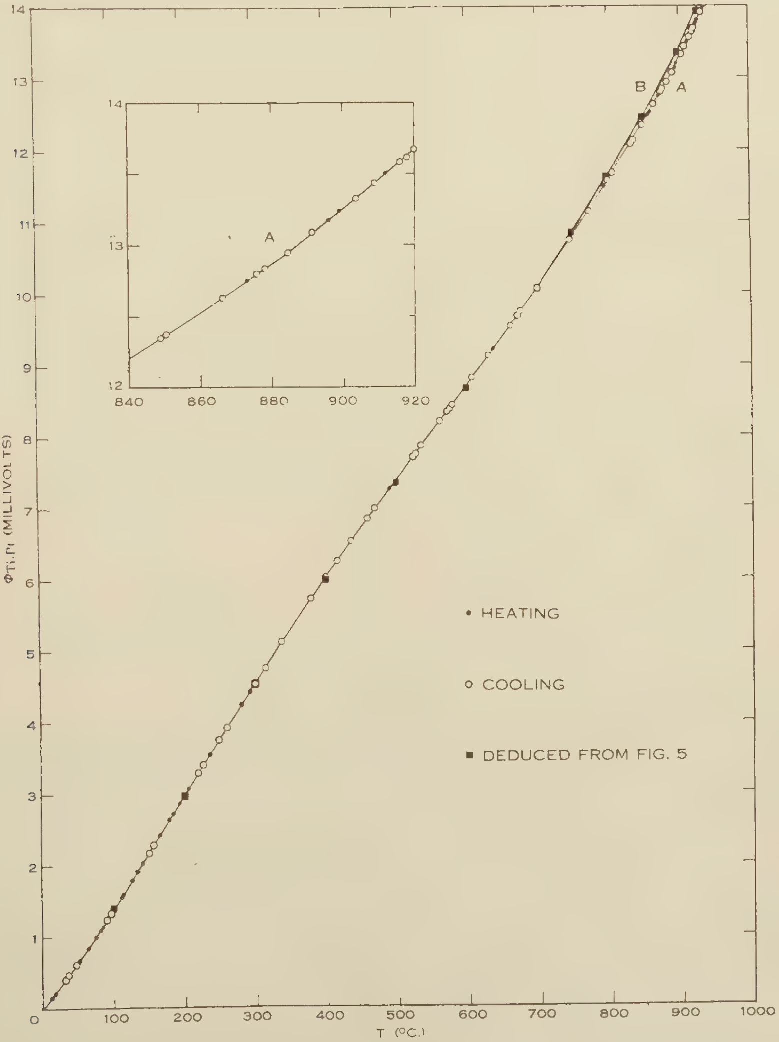


Fig. 4.—e.m.f. of titanium/platinum couple ($\varphi_{Ti/Pt}$) *versus* temperature. Cold junction temperature=0 °C.

Curve *A*. Results of direct measurement.

Curve *B*. Deduced from Figure 5 by graphical integration.

Inset. Details of curve *A* in range 850 to 920 °C.

titanium ribbon, each length having been heated *in vacuo* for ½ hour at 920 to 940 °C. The temperature of the spot-welded junction nearest the hot junction did not exceed about 350 °C. during the measurements ; hence there was no risk of the titanium near the weld becoming contaminated by dissolving the thin oxide film formed during the welding operation.

The results of e.m.f. measurements made during both heating and cooling in the range 20 to 940 °C. are collected together graphically in curve *A* of Figure 4. All the points lie on one curve. This is to be expected in view of the titanium being in the fully annealed condition with little or no preferred orientation. In other words, there was no possibility of irreversible effects coming into play during heating or cooling.

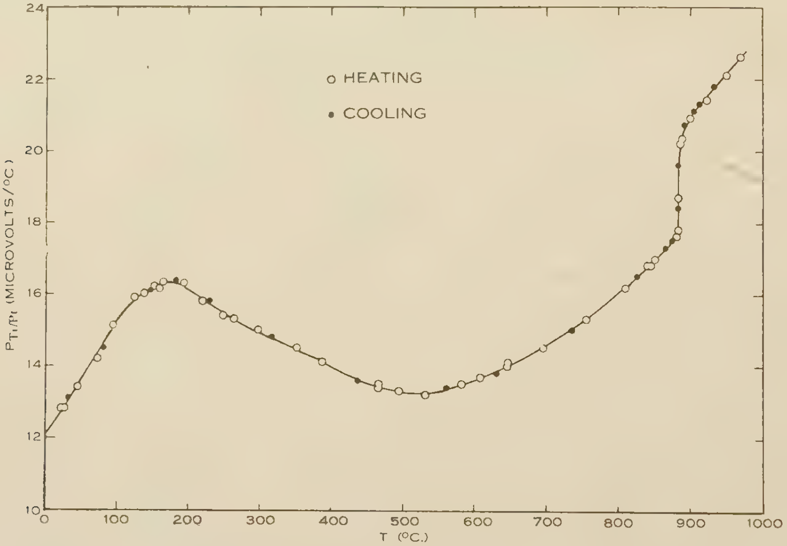


Fig. 5.—Thermoelectric power of titanium/platinum couple ($P_{Ti/Pt}$) *versus* temperature.

Careful inspection will reveal that there are two inflexions in the curve, one at approximately 200 °C., the other between 500 and 600 °C. There is a comparatively abrupt increase in gradient between 880 and 900 °C. It will be seen later that the differential results reveal these features quite markedly.

(*c*) *Results of Differential Measurements*

Figure 5 presents the results of thermoelectric power determinations made during both heating and cooling in the range 20 to 970 °C. As was the case in the direct e.m.f.-temperature determinations, there did not appear to be any evidence of irreversible effects during heating or cooling. The inflexions noted in the e.m.f.-temperature curve are indicated by the maximum at 180 °C. and the minimum at 530 °C. However, the most interesting aspect of the results shown in Figure 5 is the abrupt change in thermoelectric power between 880 and 890 °C. This is taken as a manifestation of the α–β transformation in the titanium and a later section of the paper will be concerned with a more detailed consideration of the changes in thermoelectric power accompanying the α–β transformation.

(*d*) *Integration of Thermoelectric Power Results*

Curve *B* in Figure 4 indicates the e.m.f.-temperature relationship for the titanium platinum couple as determined from the thermoelectric power results (Fig. 5) by graphical integration. The agreement between curves *A* and *B* is very satisfactory up to 800 °C. At higher temperatures, the integration results tend to be higher than the values determined by direct measurement, the difference being nearly 1 per cent. in the neighbourhood of 900 °C. In view of the electrical difficulties which are inherent in differential methods such as the one employed in this work, the agreement between the two sets of results is considered to be satisfactory.

(*e*) *Peltier and Thompson Coefficients and Absolute Thermoelectric Power*

These properties may be calculated from the thermoelectric power. The Peltier coefficient, $\pi_{Ti/Pt}$, for the titanium/platinum couple is given by

$$\pi_{Ti/Pt}=\theta \cdot P_{Ti/Pt},$$

where θ is the temperature on the absolute scale, $P_{Ti/Pt}$ being the thermoelectric power. The difference between the Thomson coefficients, σ, of the two metals is

$$\sigma_{Ti}-\sigma_{Pt}=\theta \cdot \frac{dP_{Ti/Pt}}{dT}.$$

In respect to the sign of the Thomson coefficients, the convention of the *Landolt Tables* has been adopted. The absolute thermoelectric power, *S*, is defined as

$$S=\int_0^\theta \frac{\sigma}{\theta} \cdot d\theta.$$

Using the above expression for $\sigma_{Ti}-\sigma_{Pt}$, it is evident that

$$S_{Ti}=S_{Pt}+P_{Ti/Pt}.$$

The values of S_{Pt} were taken from the data of Borelius (1943). Here it should be noted that the values of S_{Pt} above 400 °C. in Borelius's list were estimated by a process of extrapolation. Hence they can be taken as approximate only.

The variation of the above properties with temperature is displayed graphically in Figure 6. Some comments relating to these data will be found in the discussion and conclusion.

(*f*) *Allotropic Transformation Range*

A detailed examination of the thermoelectric power *versus* temperature relationship in the vicinity of the allotropic change has served to indicate the exact range over which the transformation occurs. In this connection, the results of the differential method have been analysed in the following way. When the gradient of the e.m.f. (φ) *versus* temperature (T) curve for a pair of metals varies continuously and comparatively slowly with temperature, then the ratio $\triangle\varphi/\triangle T$, that is, the ratio measured by the differential method, is, for all practical purposes, equal to $P (P = d\varphi/dT)$. Of course this will hold only if the temperature differential $\triangle T$ employed in the measurements is not more than a few degrees C. But in the immediate vicinity of any abrupt change in the gradient of the φ *versus* T relationship, the measured ratio $\triangle\varphi/\triangle T$ will not, in general, be equal to P although there will be an abrupt discontinuity in the $\triangle\varphi/\triangle T$ *versus* T curve.

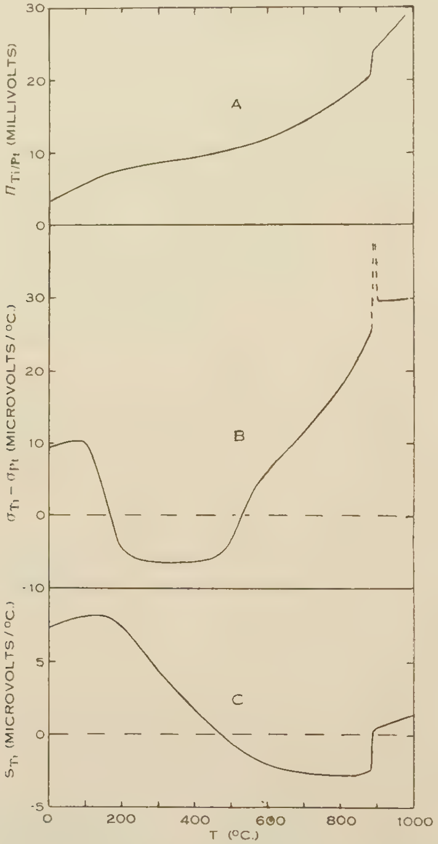


Fig. 6.—Thermoelectric coefficients and absolute thermoelectric power *versus* temperature.

Curve *A*. Peltier coefficient, $\pi_{Ti/Pt}$.
Curve *B*. Difference between Thomson coefficients of titanium and platinum, $(\sigma_{Ti} - \sigma_{Pt})$.
Curve *C*. Absolute thermoelectric power of titanium, S_{Ti}.

If the abrupt change in P occurs over a small range of temperature T' to T'' $(T'' > T')$, while the temperature differential in the method of measurement is $T_2 - T_1 = \triangle T$, the latter being of the same order as $T'' - T'$, then the range of temperature, T_l to T_h, over which there will be a comparatively abrupt change in $\triangle\varphi/\triangle T$, is given by

$$T_h - T_l = T'' - T' + \triangle T.$$

Naturally, $\Delta\varphi/\Delta T$ is plotted against the mean temperature $(T_1+T_2)/2$, T_1 and T_2 being the temperatures at the differential thermocouple junctions. Hence

$$T' = T_l + \frac{\Delta T}{2}$$

and

$$T'' = T_h - \frac{\Delta T}{2}.$$

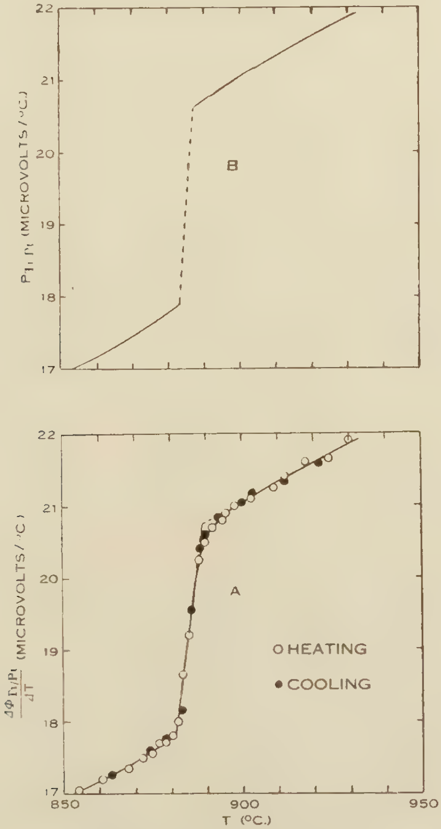


Fig. 7.—Curve *A*. Details of $(\Delta\varphi_{Ti/Pt})/\Delta T$ *versus* T relationship in the allotropic transformation range of refined titanium. ($\Delta T = 4{\cdot}2$ °C.)
Curve *B*. $P_{Ti/Pt}$ *versus* T relationship deduced from curve *A*.

In Figure 7, curve *A* indicates the $(\Delta\varphi_{Ti/Pt})/\Delta T$ *versus* T relationship for a sample of refined titanium, ΔT being 4·2 °C. Curve *B* indicates the $P_{Ti/Pt}$ *versus* T relationship deduced from curve *A* in the way outlined in general terms above. In this particular case, T_l=881 °C., T_h=889 °C., and since ΔT=4·2 °C., T'=883 °C., and T''=887 °C.

It will be noticed that whereas there is a fairly sharp change of gradient in curve *A* at 881 °C., the change in gradient is not so sharply defined at the upper end of the transformation range, that is, in the vicinity of 888 to 894 °C. The author believes that this is probably due to slight superficial contamination of the specimen with oxygen and nitrogen. The presence of either or both of these elements in titanium elevates the α–β transformation range. Therefore a thin surface layer containing a small concentration of oxygen and nitrogen would account for a rounding off of the $\triangle\varphi/\triangle T$ *versus* T curve at the upper end of the transformation range. With this idea in mind, the almost linear portion of curve A in the β range (900 to 930 °C.) was extrapolated to intersect the extrapolation of the linear portion of the curve in the transformation region. The temperature corresponding to the point of intersection (889 °C.) was used in estimating the upper limit of the true $\alpha+\beta$ range of the specimen, disregarding the thin, contaminated superficial layer. Thus curve B in Figure 7 represents the change of $P_{Ti/Pt}$ with T, approximate allowance having been made for the effects of superficial contamination. It is difficult to estimate exactly how P varies with temperature within the transformation range ; hence this portion of curve B is dotted.

In the $\alpha+\beta$ range, a steady value of $\triangle\varphi/\triangle T$ was reached within 3 to 5 minutes after the hot junction temperatures T_1 and T_2 had been established at steady values. This indicated that the transformation rate was quite high. It was therefore not surprising that there was little or no difference between the thermoelectric power-temperature relations in heating and cooling experiments. In one case, the transformation range appeared to be 2 or 3 °C. lower during cooling than it was during heating.

Table 3 presents the results of four separate determinations of the $\alpha+\beta$ range in refined titanium. The data in Table 3 can be summarized by stating that the transformation range is 883 ± 1 to 887 ± 2 °C.

TABLE 3

ALLOTROPIC TRANSFORMATION RANGE IN REFINED TITANIUM AS DETERMINED FROM THERMOELECTRIC POWER - TEMPERATURE RELATIONSHIPS

Experiment No.	Transformation Range (°C.)	
	Heating	Cooling
1 (Fig. 7)	883–887	883–887
2	884–889	884–890
3	883–889	881–886
4	882–887	883–887

If a specimen was heated and cooled three or four times in a range 850 to 950 °C. over a period of several hours, the changes in $d(\triangle\varphi/\triangle T)/dT$ at the extremities of the $\alpha+\beta$ range became less abrupt. This applied particularly to the upper limit of the range. Also, prolonged heating of the specimen tended to cause elevation of both the lower and upper limits, but more particularly the latter. These effects, which were noted even when the pressure was kept down in the range 10^{-5} to 5×10^{-5} mm. mercury, can readily be explained on the basis of a slow, progressive absorption of oxygen and nitrogen into the titanium during the course of the measurements.

All the data in Table 3 pertain to specimens which were heated and cooled for the first time after the treatment. Hence the titanium had suffered the minimum of superficial contamination of the type mentioned above.

In connection with the problem of superficial contamination, it is worthy of note that the bore of the porcelain insulators which held the titanium specimens became dark grey during the experiments. Chemical tests showed that the dark layer contained titanium. This suggested that at temperatures of the order of 900 °C. the rate of vaporization of titanium was appreciable. There was therefore the risk of titanium vapour reacting with the insulator, generating the comparatively volatile compound silicon monoxide, some of which might contaminate the titanium. When silica insulators were tried, some such process as this must have occurred, because the specimens become coated with a silicon-bearing layer. However, there was only slight evidence of this occurring when new aluminous porcelain insulators were employed and when the porcelain insulators had been used previously, there was actually no detectable absorption of silicon into the titanium surface. For this reason, it was advantageous to use one porcelain insulator repeatedly.

(*g*) *Effect of Certain Impurities on the Transformation Range*

Some effects attributed to unintentional surface contamination have already been mentioned in the previous section. Here it is proposed to present some results demonstrating quantitatively the marked influence of certain impurities on the transformation range in titanium. In Figure 8, curve *A* indicates the variation of thermoelectric power with temperature in a sample of refined titanium. Curve *B* indicates the effect of adding approximately 0·9 at.–% oxygen to the refined titanium. It is evident that the 0·9 at.–% oxygen has elevated and broadened the transformation range to 898 to 920 °C. Curve *C* in Figure 8 pertains to titanium of commercial purity (cf. Table 1). In this grade of metal, the transformation occurs over a very broad range of temperature, namely, 865 to 970 °C. The extent of this interval is somewhat surprising, especially in respect to the fact that the lower limit of the transformation range (865 °C.) is lower than that in the refined titanium. This might be attributable to the presence of iron which is known to depress the $\alpha-\beta$ range in titanium. Of the other major impurities, oxygen and nitrogen are known to elevate the transformation. Between 910 and 950 °C., the gradient of curve *C* is lower than elsewhere within the transformation range. In view of the number and content of the impurities in the commercial grade of metal, it is difficult to interpret this change of gradient in curve *C*.

(*h*) *Some Effects of Cold Work, Preferred Orientation, and Annealing*

All the results presented heretofore relate to annealed titanium specimens in which there was no appreciable degree of preferred orientation. It was considered advisable to conduct some experiments which would reveal the magnitude of effects due to cold work and preferred orientation. Of course it was impossible to examine these two factors independently since cold working always tends to produce some degree of preferred orientation.

The stock of refined titanium ribbon had been produced by cold rolling and the reduction in thickness subsequent to the intermediate annealing treatment was approximately 80 per cent. By means of X-ray diffraction experiments, it was revealed that this cold rolled ribbon exhibited preferred orientation, the (0001) planes tending to lie parallel to the plane of rolling.

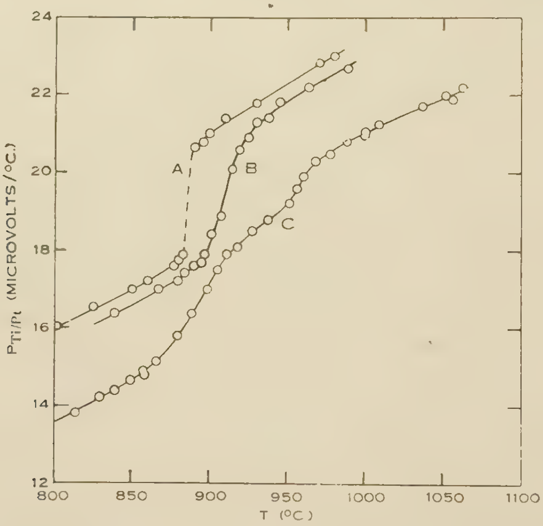


Fig. 8.—Effect of impurities on α—β transformation range of titanium.

Curve *A*. Refined titanium.
Curve *B*. Titanium containing 0·9 at.-% oxygen.
Curve *C*. Commercial grade of titanium.

A sample of the " as rolled " ribbon was employed in a slow heating experiment, the results of which are presented in Figure 9, curve *B*. Curve *A* in Figure 9, pertaining to an annealed specimen with no preferred orientation, is included so that comparisons can be made readily. At temperatures up to 330 °C., curve *B* lies above curve *A*, while the reverse is true from 330 °C. up to the commencement of the transformation zone. Within the transformation range, curve *B* rises steeply so that it becomes almost coincident with curve *A* in the β range above 890 °C.

Here it is worthy of note that when the specimen used to obtain curve *B* was cooled back into the α region, curve *B* was not retraced ; actually curve *A* was followed very closely. This suggested that heating into the β range, followed by cooling back into the α range, eliminated preferred orientation. This observation was in concordance with the results of X-ray diffraction experiments, mention of which has already been made in Section III (*a*).

Reverting to the heating from room temperature up to about 300 °C., it was noted that steady values were not attained in the range 130 to 280 °C. This is shown in Figure 9 by the arrows which are meant to indicate that on first heating the specimen to a particular temperature, the thermoelectric power was high, but tended to fall, at first rapidly, then slowly. Hence the values to which the arrows point are not equilibrium values; actually they represent results of measurements made after about 2 hours at each temperature. The observations just recorded prompted an exploratory study of the annealing of the cold rolled titanium strip. The annealing experiments, results of which are set out in Figure 10, revealed that the first definite signs of softening occurred on heating to 190 °C. for 1 hour. Thus it is not unlikely that the instability of thermoelectric power in the range 130 to 280 °C. was connected with the recovery stage in the annealing of the cold rolled titanium.

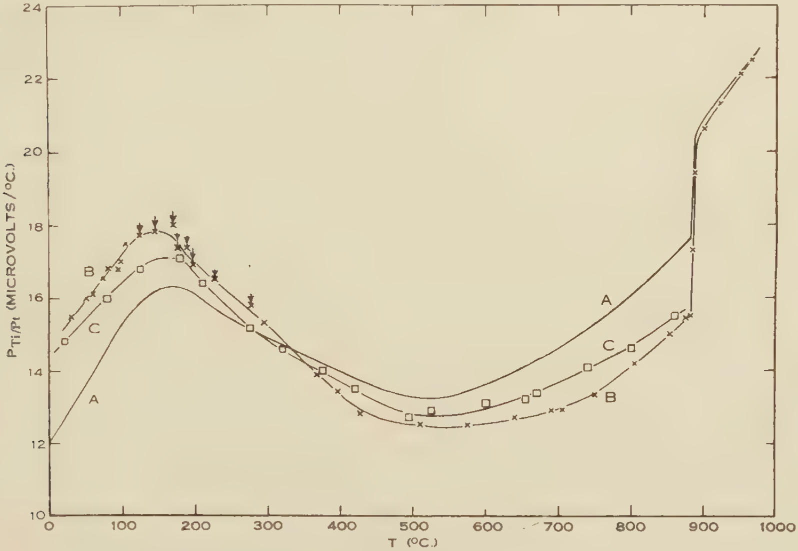


Fig. 9.—Thermoelectric power *versus* temperature relationships showing effects of preferred orientation and annealing.

Curve *A*. Annealed titanium with no preferred orientation.

Curve *B*. Specimen initially in cold rolled condition. Heating experiment.

Curve *C*. Titanium strip annealed for 1 hour at 860 °C. Heating experiment.

The results of the annealing experiments (Fig. 10) indicated that heating for about 1 hour at 860 °C. would cause complete softening and recrystallization of the titanium strip. However, X-ray diffraction experiments showed that there was still an appreciable degree of preferred orientation in the strip after the above annealing treatment, the type of orientation being the same as in cold rolled strip. Therefore the thermoelectric properties of the titanium after heating for an hour at 860 °C. are likely to be of interest in revealing the effect of preferred orientation without the complications of recovery and recrystallization coming into play. In this connection, an experiment was conducted and the results are presented as curve *C* in Figure 9. This curve lies between curves *A* and *B*, except in the range 300 to 350 °C. It is clear that the anisotropy of thermoelectric power in α titanium is quite marked except in the vicinity of 250 to 300 °C. where curves *A* and *C* cross.

After the results shown in curve *C*, Figure 9, were obtained, the specimen was heated through the $\alpha-\beta$ range. During heating through the transformation region, the thermoelectric power of the couple rose steeply until at 890 to 900 °C., the values fell very close to curves *A* and *B*. On subsequent cooling back into the α region, curve *C* was not re-traced ; rather was the course of *A* followed. These observations indicate more or less complete obliteration of the preferred orientation as a result of heating into the β zone (cf. discussion above in relation to curve *B*).

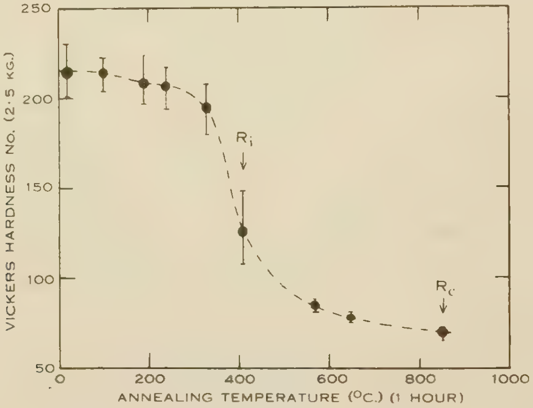


Fig. 10.—Results of annealing experiments on refined titanium. Initial state of metal—cold rolled, 80 per cent. reduction in thickness subsequent to previous annealing at 750 °C.

R_1 indicates stage at which recrystallization was first noted ; R_c denotes complete recrystallization.

IV. Discussion and Conclusion

The results of the present investigation have demonstrated the usefulness of the differential method for determining thermoelectric properties. In particular, the method has yielded results of value concerning the allotropic transformation range in titanium refined by the iodide process. The range as determined from the thermoelectric power results is 883 ± 1 to 887 ± 2 °C. These figures are in reasonable agreement with the data of McQuillan (1950), who indicated that the $\alpha \rightarrow \beta$ change in refined titanium commenced at $882 \cdot 5 \pm 1$ °C. McQuillan employed two properties, namely, electrical resistance and hydrogen solubility, and his results showed that the upper limit of the $\alpha+\beta$ range was less sharply defined than the lower limit. The same observation was made in the present investigation.

It has been shown that the impurities present in a commercial grade of titanium broaden the transformation range out to 865 to 970 °C. The major portion of this effect seems to be due to the non-metallic impurities, oxygen, nitrogen, and carbon. The marked effects of preferred orientation have been revealed. There are indications that the preferred orientation is practically eliminated on heating just above the $\alpha + \beta$ range, and in this connection the thermoelectric power results and X-ray diffraction observations are in concordance with one another.

The results of the differential measurements have been used to establish the temperature dependency of properties of more theoretical interest, namely, (*a*) the Peltier coefficient of the titanium/platinum couple, (*b*) the difference between the Thomson coefficients of titanium and platinum, and (*c*) the absolute thermoelectric power of titanium. From the theoretical standpoint, the latter is probably of greatest interest. However, it appears that application of the electron theory of metals to thermoelectricity cannot as yet provide a definite explanation of the effects shown in Figure 6. The theoretical treatment of thermoelectric effects by Mott and Jones (1936) would suggest that the absolute thermoelectric power of transition metals would be negative at normal temperature. Actually, the value for titanium at 20 °C. is $+7 \cdot 5$ microvolts/°C. It is evident that the absolute thermoelectric power of β titanium at, or near the transformation temperature region is higher than that of the α form. This provides an example of what appears to be a more or less general tendency, namely, that the absolute thermoelectric power of a close-packed form of a metal is lower than that of a form in which the atomic packing is not " close ", the comparison being made at the temperature at which the two forms are in equilibrium. The rule seems to apply to allotropic transformations (e.g. iron and titanium) and to the fusion or solidification of metals. In the latter case, the liquid is considered as the more " open " structure.

V. ACKNOWLEDGMENTS

The author gratefully acknowledges the use of facilities kindly provided by Professor J. Neill Greenwood. Mr. J. Corbett very ably performed practically all of the chemical analyses and the spectrographic testing was carried out by the Defence Research Laboratories, Maribyrnong, Victoria.

VI. REFERENCES

BARBER, C. R. (1950).—*Proc. Phys. Soc.* B **63** : 492.

DE BOER, J. H., BURGERS, W. G., and FAST, J. D. (1936).—*Proc. Acad. Sci. Amst.* **39** : 515.

BORELIUS, G. (1943).—" Handbuch der Metallphysik." Vol. 1, p. 398. (Akademische Verlagsgesellschaft M.B.H. : Leipzig.)

GREINER, E. S., and ELLIS, W. C. (1949).—*Trans. Amer. Inst. Min.* (*Metall. Engrs.*) **180** : 657.

McQUILLAN, A. D. (1950).—*J. Inst. Met.* **78** : 249.

MOTT, N. F., and JONES, H. (1936).—" The Theory of the Properties of Metals and Alloys." p. 313. (Oxford Univ. Press.)

SQUIRE, C. F., and KAUFMANN, A. R. (1941).—*J. Chem. Phys.* **9** : 673.

WORNER, H. W. (1949).—*Metallurgia Manchr.* **40** : 69.

AQUEOUS DIFFUSION MEASUREMENTS BY THE LAMM SCALE METHOD AND THEIR CORRELATION WITH MEMBRANE DIFFUSION MEASUREMENTS

By D. G. Gillam*

[*Manuscript received September* 20, 1950]

Summary

The diffusion of potassium chloride, sodium chloride, potassium sulphate, and sucrose in water is investigated by means of a simplified form of the Lamm scale method. The maximum height method of calculation is modified in such a way that the experimental procedure is less laborious than the original method. The diffusion coefficient values are obtained over a range of temperatures and the activation energies for free diffusion calculated. Comparisons are made with the predictions of the Debye-Hückel-Onsager theory and with the results of membrane diffusion measurements. The defects of the method are discussed and a more refined method is described by which it should be possible to calculate the concentration dependence of the diffusion coefficient and also the influence of solvent counterflow.

I. Introduction

Diffusion rates for electrolytes through membranes (Tolliday, Woods, and Hartung 1949) cannot be correlated with those for free diffusion in water because of the discordance of the published values of the latter. In a general investigation of diffusion it is desirable to determine the concentration c or the concentration gradient dc/dx as a function of the time t throughout the whole of the diffusion system. This is so difficult and time-consuming that simpler methods are adopted. Recently, diffusion has been investigated mainly by the conductimetric measurement of concentration or by the optical measurement of concentration gradient. The main disadvantages of the former are that the presence and operation of electrodes in the diffusion column may have some effect on the gradients and that measurements cannot be limited to definite levels because of the finite size of the electrodes. The measurements themselves are, however, very accurate over a wide range of concentrations and, in dilute solutions, may be used to provide a check on the correctness of the Debye-Hückel-Onsager theory. On the other hand, optical methods enable the determination of dc/dx at very closely defined levels in the column, and the diffusion process is not disturbed in any way. However, accurate refractive index data at small concentration intervals are required for evaluating the concentration gradients, and published information is, in general, not sufficiently reliable for this. Further work on the determination of refractive indices is therefore needed before optical methods can be developed much further. These methods depend basically on

---

* Chemistry Department, University of Melbourne.

a knowledge of the gradient of refractive index with concentration and any method for determining this gradient dn/dc directly (e.g. by the interferometer) would be ideal. In general, all optical methods are derived from two fundamental methods described by Lamm (1937), and these have been developed to give a direct registration of the dn/dx against x curve. However, the optical systems required for this are complicated and the simpler original methods may be used without sacrifice of accuracy.

If the variation of dc/dx with time for all levels in the column can be determined, then the variation of diffusion coefficient with concentration may be calculated as in the method developed by Calvet and Chevalerias (1946). In the Lamm (1937) scale method, the angular deviations of thin beams of light passed through the diffusion column are determined. Since the deviation is proportional to the refractive index gradient at the point through which the light beam passes, accurate values for this gradient can be obtained at known heights throughout the column. The magnitudes of the deviations are determined by focusing the resultant light beams by means of a lens and measuring the displacements at the focal plane. The Lamm scale method, hence, looked promising for the investigation of rates of diffusion in water. However, neither this method nor that of Calvet and Chevalerias can determine the influence of solvent counterflow which, as Clack (1908, 1912, 1914, 1916, 1921, 1924, 1925) has pointed out, is far from negligible in liquid diffusion. Solvent counterflow is the term used to describe the " mutual diffusion " which occurs in the liquid phase. Each component in a diffusing system will move against a stream of the other component moving in the opposite direction. Most workers in the aqueous diffusion field have measured diffusion coefficients relative to a frame of reference given by the walls of the containing vessel. However, since the solvent is moving during the diffusion process with variable speed in different sections of the diffusion column, this frame of reference is not altogether satisfactory. Clack preferred to use a frame of reference fixed with respect to the solvent and developed a steady state method with determinable velocity of solvent counterflow. He was thus able to obtain diffusion coefficients referred to an ideal system in which the solvent was at rest. Clack's method is the only one, so far devised, which takes into account the influence of solvent counterflow. The theoretical calculations of Onsager (1932, 1945) also refer to the stationary solvent frame of reference.

The Lamm scale method was selected for investigation with the object of improving it. In Wiener's (1893) solution of Fick's law, for use in optical methods for the sharp initial boundary case, refractive index is assumed proportional to concentration. This assumption is incorrect and the following solution, for the concentration gradient in the cell, is preferable

$$\frac{dc}{dx}=\frac{c_1-c_0}{2\sqrt{\pi Dt}}e^{-x^2/4Dt}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (1)$$

This follows directly from Fick's law where c_1, c_0 are the concentrations on either side of the sharp boundary at time $t=0$.

The maximum value of dc/dx always remains at the level of this boundary ($x=0$) if D is constant, that is,

$$\left(\frac{dc}{dx}\right)_{max}=\frac{c_1-c_0}{2\sqrt{\pi Dt}}. \quad \ldots\ldots\ldots\ldots \quad (2)$$

The "maximum height" method (Lamm 1937) for the calculation of the diffusion coefficient is based on this equation.

Now

$$\left(\frac{dn}{dx}\right)_{max}=\left(\frac{dc}{dx}\right)_{max}\cdot\left(\frac{dn}{dc}\right)_{c_A}, \quad \ldots\ldots\ldots\ldots \quad (3)$$

where $(dn/dx)_{max}$ is the maximum gradient of refractive index in the cell and $(dn/dc)_{c_A}$ is the rate of change of refractive index with concentration at the concentration c_A at which the maximum displacement occurs. This concentration equals $(c_1+c_0)/2$ if Fick's law is true.

Hence

$$D=\frac{(c_1-c_0)^2\left(\frac{dn}{dc}\right)^2_{c_A}}{4\pi t\left(\frac{dn}{dx}\right)^2_{max}}, \quad \ldots\ldots\ldots\ldots \quad (4)$$

where $(dn/dc)_{c_A}$ can be determined from refractive index-concentration tables and $(dn/dx)_{max}$ from experimental measurements using the image-displacement equation

$$Z_{max}=G\,a\,b\left(\frac{dn}{dx}\right)_{max},$$

where Z_{max}=maximum scale line displacement produced on the photographic plate (cm.),

G=optical magnification,

a=thickness of diffusion column in direction of light beam (cm.),

b=optical scale distance (cm.).

Hence

$$D=\frac{(c_1-c_0)^2G^2a^2b^2(dn/dc)^2_{c_A}}{4\pi tZ^2_{max}}. \quad \ldots\ldots\ldots\ldots \quad (5)$$

In any set of experiments G, a, b, t, c_1, c_0 remain constant and Z_{max} is found by measuring the displacement of a few lines on a photographic plate by means of a comparator. This method is therefore much less laborious then the original Lamm procedure involving accurate measurement of several hundred lines.

II. Experimental

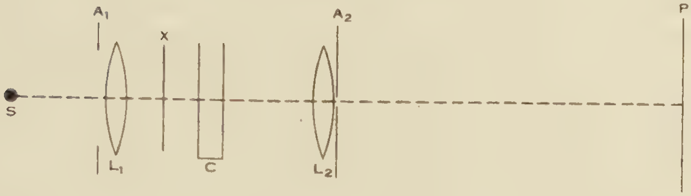


Fig. 1

X is a vertical scale with a line spacing of 0·05 cm. and line thickness of 0·008 cm. C is a rectangular glass diffusion cell, 10 mm. thick, with plane parallel walls. The distance of the scale from the middle of the diffusion cell (the scale distance) is 12 cm. The optical scale distance, b, equals the sum of the geometrical distances involved in the scale distance, divided in each case by the corresponding refractive indices of the media. The scale is illuminated by the light from the source S by means of the condensing lens L_1. The lens L_2, of focal length 40 cm., was stopped down to $f : 60$ by means of the aperture stop A_2. A specially designed holder was constructed for the photographic plate P to record the image of the scale so that either vertical half of the plate could be exposed independently. The middle of the scale line image falls on the dividing line between the two halves of the plate. In this way, images of the undisplaced and the displaced line series can be obtained in juxtaposition. The displacements are determined subsequently by a comparator.

The whole optical system was placed in an air thermostat consisting of a long double-walled box, provided internally with suitable electrical heaters and circulating fans, controlled by a mercury-toluene regulator with thyratron relay. In this way the internal temperature could be controlled to ±0·2 °C. for long periods. Preliminary experiments showed that there was no difference in the diffusion coefficients obtained when the temperature fluctuations were ±0·5 and ±0·1 °C., proving that control to ±0·2 °C. was adequate for this work. It was demonstrated that an initial temperature equilibration period of 3 to 4 hours for the thermostat was required when heating the apparatus from room temperature to the required experimental temperature. It was also shown that there was no alteration in the value of the diffusion coefficient when, instead of the above procedure, the air temperature outside the thermostat was automatically controlled.

To avoid vibration, the thermostat was mounted on a heavy slate slab supported by three concrete pillars let into the ground in the basement of the laboratory. Since no improvement was found when the optical bench was supported in addition in the thermostat on a heavy spring-loaded platform, vibration errors were regarded as negligible.

The initial boundary was formed as follows : A sample of the solution, from a reservoir which was situated in the thermostat, was run into the diffusion cell. Then a finely drawn out capillary tube, which was connected to a reservoir containing water, was touched against a wall of the diffusion cell just above the surface of the solution. The water in the capillary was then allowed to run out very slowly and it thus formed, in time, a layer on top of the solution. After the required amount of water had been run in, the capillary was removed from the diffusion cell. The internal diameter of the capillary was 0·027 cm. and the volume rate of outflow through the capillary was 0·008 cc./sec. All these operations, including the boundary formation, were carried out mechanically in order that the temperature equilibrium inside the thermostat would not be disturbed. The starting time was taken when the first drop of water ran out of the capillary tube. This method is simple, the starting time is definite, the boundary is initially sharp and its position in the diffusion cell is not shifted during the experiment.

The image-displacement equation $Z=G\,a\,b\dfrac{dn}{dx}$ used in the scale method has been derived by Lamm (1937) from classical optics on the assumptions :

(i) All angles involved are so small that their sines and tangents may be replaced by the angles measured in radians.
(ii) The refractive index n is constant along the light path in the cell.
(iii) The refractive index gradient dn/dx is constant along the light path in the cell.

It was calculated that all the angles involved would be less than 0·006 radian and hence the first assumption was correct. The Lamm (1928, 1937) theory was used to calculate the maximum variation of refractive index (5×10^{-5}) and of the refractive index gradient (2×10^{-5} cm.$^{-1}$) along the light path. These values are negligible compared with the refractive indices (*c.* 1·3) and the refractive index gradients (*c.* 6×10^{-3} cm.$^{-1}$) measured in the column.

It was necessary also to investigate possible temperature alterations which could be produced in the column itself by the heat of dilution. The heat of dilution is intimately bound up with the diffusion process, and a formal separation of the two processes does not appear possible. However, the order of magnitude of the effect can be calculated. Aqueous diffusion is a very slow process and any local heat effects are quickly nullified by conduction in the liquid itself, by heat transfer to the walls of the containing vessel and by convection. This last is the most serious effect in diffusion work. A simple calculation showed that this temperature effect is quite small ; for potassium chloride in water it cannot exceed 0·04 °C. and as dD/dT for most substances is about 3 per cent. per degree, this value for the temperature change would introduce an error of 0·1 per cent., which is well within the experimental errors of measurement.

The maximum error in the diffusion coefficient value calculated from equation (5), due to errors in the physical measurements, is estimated to be ±2 per cent. The method described is reasonably simple and rapid because the number of lines, which need to be measured by the comparator, is small and the diffusion column is not disturbed by the method of measurement. The accuracy is limited, however, by the quality of the standard refractive index tables for solutions and the difficulty of applying a correction for the influence of solvent counterflow.

III. RESULTS

Results were obtained for the diffusion of potassium chloride, sodium chloride, potassium sulphate, and sucrose at various temperatures. It was found that any measurement could be duplicated to 2 per cent. Each value in Table 1 is the mean of two or three determinations.

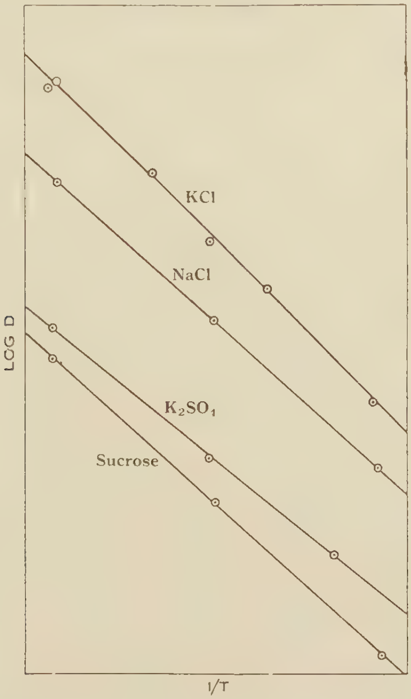


Fig. 2

TABLE 1

DIFFUSION COEFFICIENT VALUES

	Sucrose			KCl		KCl		NaCl		K_2SO_4	
c_1 (g./l.)	20			60		40		40		60	
t (hr.)	1	2	4	2		2		2		2·5	
T^1 (°C.)	D (cm.²/sec.) $\times 10^5$			T^1	$D \times 10^5$	T^1	$D \times 10^5$	T^1	$D \times 10^5$	T^1	$D \times 10^5$
19·0		$0{\cdot}46_1$		19·2	$1{\cdot}43_5$	21·0	$1{\cdot}51_5$	19·1	1·21	19·9	$1{\cdot}03_5$
22·0	$0{\cdot}50_3$	$0{\cdot}49_7$	$0{\cdot}49_8$	22·1	1·55	23·1	$1{\cdot}60_5$	22·0	1·30	22·1	$1{\cdot}08_5$
25·0		$0{\cdot}53_2$		25·1	1·67	24·9	$1{\cdot}67_5$	25·1	1·39	25·0	$1{\cdot}15_5$

Figure 2 shows the straight lines obtained by plotting log *D* against 1/*T* (vertical scale arbitrary), the values for both potassium chloride solutions falling on the same straight line.

The activation energies for the diffusion process, obtained from the slopes of the lines, are compared in Table 2 with the values obtained from the Nernst equation and with those obtained for diffusion through cupric ferrocyanide membranes (Tolliday, Woods, and Hartung 1949).

TABLE 2

Electrolyte	Activation Energy (cal./mol.) in 19–25 °C. Range		
	Free Diffusion		Membrane Diffusion
	Nernst Equation	Present Investigation	
KCl ..	4,100	4,400	5,100
NaCl ..	4,300	4,100	
K_2SO_4 ..	4,200	3,700	5,900
Non-electrolyte : sucrose ..		4,100	

In addition, activation energies, at definite values of the concentration, were calculated by means of the Debye-Hückel-Onsager theory (Harned and Owen 1943) for potassium chloride and the results are included in Table 3.

TABLE 3

c (moles/litre)	Activation Energy	Source
0	4,100	Nernst equation
0·06	4,100	Onsager theory
0·25	4,100	Öholm's (1905, 1910) investigation
0·25	4,200	Onsager theory
1·0	4,300	Öholm's (1905, 1910) investigation
1·0	4,400	Present investigation
1·0	4,400	Onsager theory
2·25	4,700	Onsager theory
4·0	5,100	Onsager theory

IV. Discussion

Within the experimental error, the value of the diffusion coefficient for sucrose did not depend on the length of the diffusion period. There are two influences which may account for this result. The first is the assumption that Fick's law is true with D = constant and the second that an infinitely sharp boundary was formed at $t=0$. Both of these factors together with use of the maximum height method of calculation may account for this inability to show any variation of the diffusion coefficient. The values for sucrose compare fairly well with those obtained by Lamm (1928). He used a concentration $c_1=2$ per cent. and diffusion periods of 1·5, 3, 4, 5 hours and found that the diffusion coefficient did not depend on the length of the diffusion period, the average value being 0·462 at 20 °C. The sucrose result obtained at 25 °C. in the present work is 3 per cent. higher than that obtained recently by Gosting and Morris (1949).

The diffusion coefficients for potassium chloride are much lower than those obtained previously by Öholm (1905, 1910), Lamm (1937), and others. It is unusual to obtain low results by a free diffusion method since such methods tend to give high values for the diffusion coefficient due to the initial boundary formation not being ideally sharp. Öholm used a layer analysis method in which the average concentrations of several layers of the diffusion column were determined analytically at the end of the diffusion period. However, although disturbances may be small at the initial boundary formation due to the compensating effect of the high density gradient, they may be large at the end of the diffusion experiment when the layers are divided up since the density gradients will be then relatively small. This would contribute to an increase in the value of the diffusion coefficient. The above results are 12 per cent. lower than the value obtained by Lamm with $c_1=0 \cdot 1$N, using the removable partition method of boundary formation and diffusion periods of 0·5, 1, 1·5, 2, 2·5 hours. This method, however, does not appear capable of producing a sharp boundary because of the finite thickness of the partition. Moreover, the density difference between the two layers in Lamm's experiments seems too small to counteract the mixing produced during boundary formation. One would expect the Lamm method to work better for sucrose than for an electrolyte, as the concentration variation of the diffusion coefficient is much smaller. The diffusion coefficients determined for sodium chloride are close to the published values. This is in contradistinction to the situation for potassium chloride and is difficult to understand as the concentration variation of the diffusion coefficient for sodium chloride is probably as great as for potassium chloride. In the case of potassium sulphate, there were insufficient values tabulated in the literature to make a comparison.

Table 2 shows that the activation energies for free diffusion are less than those for membrane diffusion, the increase of the potential barrier being due to the effect of the pore size and pore wall charges of the membranes.

Because of the dependence of the diffusion coefficient on concentration and of the influence of solvent counterflow, it is evident that the above method should be developed so that both of these effects can be investigated. At present, Clack's (1908, 1912, 1914, 1916, 1921, 1924, 1925) method is the only one by which the solvent counterflow correction can be calculated. In his method, the steady state of diffusion is set up in a rectangular glass cell by passing a stream of water across the top of the cell and by sealing, to the lower end, a reservoir containing a saturated solution plus excess solid. This method removes the uncertainty of the sharp initial boundary formation, but Clack's results indicate that the stream of water disturbs the gradients in the steady state cell. One way of overcoming this defect is to seal off the ends of the column by means of sintered glass discs. This would mean that the steady state conditions could be accurately maintained at the surfaces of the discs without disturbing the diffusion column itself. The concentration gradients in the cell would be measured by an optical method similar to the one described above. In the steady state case, however, one is permitted to use cells which are much wider than in the diffusion from a sharp initial boundary. This is due to the fact that the concentration gradient is almost constant along the steady state column and, hence, the refractive index variation is the deciding factor on the width of the cell. Since this variation is small, cells of 4 cm. width can be used. In the other case, the refractive index gradient varies quite rapidly with height in the cell and the maximum permissible cell width is 1·5 cm. With the wider cell, greater displacements are obtained on the photographic plate and hence more accurate measurements of the concentration gradients can be made. It is now proposed to investigate diffusion by this modification of Clack's steady state cell.

V. ACKNOWLEDGMENT

The author wishes to express his indebtedness to Professor E. J. Hartung for suggesting this research and for his help and advice during the investigation.

VI. REFERENCES

BRIDGMAN, W. B., and WILLIAMS, J. W. (1942).—*Ann. N.Y. Acad. Sci.* **43** : 195.

CALVET, E., and CHEVALERIAS, R. (1946).—*J. chim. Phys.* **43** : 37.

CLACK, B. W. (1908).—*Proc. R. Phys. Soc. Lond.* **21** : 374.

CLACK, B. W. (1912).—*Proc. R. Phys. Soc. Lond.* **24** : 40.

CLACK, B. W. (1914).—*Proc. R. Phys. Soc. Lond.* **27** : 56.

CLACK, B. W. (1916).—*Proc. R. Phys. Soc. Lond.* **29** : 49.

CLACK, B. W. (1921).—*Proc. R. Phys. Soc. Lond.* **33** : 259.

CLACK, B. W. (1924).—*Proc. R. Phys. Soc. Lond.* **36** : 313.

CLACK, B. W. (1925).—*Proc. R. Phys. Soc. Lond.* **37** : 116.

GOSTING, L. J., and MORRIS, M. S. (1949).—*J. Amer. Chem. Soc.* **71** : 1998.

HARNED, H. S., and OWEN, B. B. (1943).—" The Physical Chemistry of Electrolytic Solutions." (American Chemical Society, Monograph 95 : New York.)

LAMM, O. (1928).—*Z. phys. Chem.* A **138** : 313.

LAMM, O. (1937).—*Nova Acta Soc. Sci. Upsal.* **10** (6) : 115 pp.

ÖHOLM, L. W. (1905).—*Z. phys. Chem.* **50** : 309.

ÖHOLM, L. W. (1910).—*Z. phys. Chem.* **70** : 378.

ONSAGER, L. (1932).—*J. Phys. Chem.* **36** : 2689.

ONSAGER, L. (1945).—*Ann. N.Y. Acad. Sci.* **46** : 241.

TOLLIDAY, J. D., WOODS, E. F., and HARTUNG, E. J. (1949).—*Trans. Faraday Soc.* **45** : 148.

WIENER, O. (1893).—*Ann. Phys. Lpz.* **49** : 105.

SULPHONAMIDES

II. STRUCTURE AND TAUTOMERISM OF SULPHAPYRIDINE, SULPHATHIAZOLE, AND SULPHANILYLBENZAMIDINE

By S. J. Angyal* and W. K. Warburton*

[*Manuscript received July 24, 1950*]

Summary

By the study of ultraviolet absorption spectra and comparison with methyl derivatives of known structure it is shown that the two forms of arylsulphonylbenzamidines described by Barber (1943) are tautomers, the more stable one having structure I.

In dilute solution, sulphathiazole is found to be mainly X, and sulphapyridine a mixture of XIII and XIV in a ratio which depends on the solvent and can be influenced by changes on the *p*-amino group. Bands of the spectra are assigned to structural features of these molecules.

I. Introduction

Nearly all therapeutically active sulphonamides are open-chain or cyclic amidine derivatives, for which two tautomeric forms, p-$NH_2.C_6H_4.SO_2.NH.\dot{C} : N.$ and p-$NH_2.C_6H_4.SO_2.N : \dot{C}.NH.$, are possible. The question which of these represents the more stable structure has been discussed in some cases, e.g. sulphapyridine and sulphathiazole, whereas in others, e.g. sulphaguanidine, no studies have been reported. Northey (1948) after summarizing the available evidence, reached this conclusion : " It appears that the question of structure, or specifically the distribution of tautomeric and resonance structures, of the heterocyclic derivatives has not been settled as yet. The possible correlations between structures and chemotherapeutic activities in this series of compounds make further study of this subject desirable." At present, the subject is still controversial. The present paper is therefore intended as a contribution to the question of structure and tautomerism.

The theoretical aspects of the amidine tautomerism have, apparently, not yet been discussed. In unsubstituted amidines the two tautomeric forms are equal but in substituted amidines the equilibrium will depend on the nature of the substituent. In the system, $NH_2\text{-}CR{:}NX \rightleftarrows NH{:}CR\text{-}NHX$, if X is an electron-attracting group, such as sulphonyl, the first form will be favoured because the group X can assume a greater share of the electrons than in the latter form. The position will be similar to the tautomeric system in amides, $NH_2\text{-}CR{:}O \rightleftarrows NH{:}CR\text{-}OH$, in which the oxygen is more electron-attracting than nitrogen and therefore the first form is more stable ; the free energy difference between the two tautomers has been estimated at 10 kcal./mole (Branch and

---

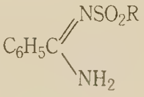
* Chemistry Department, University of Sydney.

Calvin 1941). The sulphonyl group has a high electron affinity and therefore the sulphanilimido group (*p*-$NH_2.C_6H_4.SO_2N$:) must be more electron-attracting than an unsubstituted imino group. No quantitative data are available for the comparison of the sulphanilimido group with either oxygen or nitrogen, but some qualitative conclusions can be drawn from the reaction of saccharin with diazomethane (Arndt and Scholz 1934) which gives both the *N*-methyl and the *O*-methyl derivatives. In this compound there is a tautomeric system in which an oxygen and a benzenesulphonimido group compete for the proton; the presence of both tautomers indicates that the electron-attraction of these groups is not very different. By contrast, phthalimide is methylated only on nitrogen (Arndt and Scholz 1934).

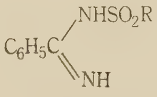
It seems reasonable therefore to assume that the more stable structure of sulphanilylamidines will be similar to that of the corresponding compounds in which an oxygen is substituted for the sulphanilimido group. By comparison with amides and urea, the stable form of sulphanilylamidines should be $NH_2.C_6H_4.SO_2.N{:}CR.NH_2$, and of sulphaguanidine, $NH_2.C_6H_4.SO_2.N{:}C(NH_2)_2$. For the latter, the same structure has been suggested, but not proved, by several authors (Jensen and Falkenberg 1942; Backer and Moed 1947; Kumler and Daniels 1943). For the 2-hydroxypyridine-pyridone and 2-hydroxythiazole-thiazolone systems the latter structures are generally accepted; sulphapyridine and sulphathiazole therefore should have similar structures, with hydrogen on the ring-nitrogens. The experiments described in this paper indicate that these conclusions are justified.

II. SULPHONYLBENZAMIDINES

The structure of some sulphonyl derivatives of benzamidine has been investigated by Barber (1943). In three cases (*p*-nitro-, *p*-acetamido-, and *p*-aminobenzenesulphonylbenzamidine) he obtained two forms, one lower melting, unstable, and one higher melting, stable, designated as α and β, respectively. Barber believed these to be the two tautomeric forms and assigned the structure I to the α-, and II to the β-forms ($R=NO_2.C_6H_4$, $AcNH.C_6H_4$, or $NH_2.C_6H_4$).



(I)

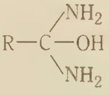


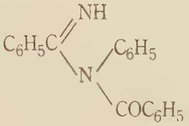
(II)

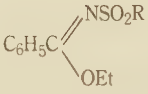
In support of this view he stated (without giving experimental details) that hydrolysis of the β-form gave only the mixed amide, $RSO_2.NH.CO.C_6H_5$, while the α-form gave some sulphonamide too.

However, Northey, Pierce, and Kertesz (1942), who obtained only the higher melting forms, assigned to them the structure I, mainly because the compounds did not form alkali salts, from which they concluded that there was no hydrogen on the sulphonamide group. On hydrolysis they obtained sulphonamides.

None of these arguments is valid. Proof of structure by hydrolysis is based on the mistaken notion that it is always the double bond which is broken. But if the hydrolysis is carried out by acids, the amidinium ion is formed first in which both nitrogen atoms have single bonds. Alkaline hydrolysis will evidently occur, as Markwald (1895) has already suggested, by addition of OH to the central carbon, followed by addition of a proton to nitrogen to give an intermediate (III) in which again both nitrogen atoms have single bonds. The products of hydrolysis will depend only on the nature of the substituent groups: thus Wheeler, Johnson, and McFarland (1903) found that *N*-phenyl-*N*-benzoylbenzamidine (IV) gave benzamide and benzanilide, that is, the single C-N bond had opened.


(III)


(IV)


(V)

The argument about the formation of alkali salts, used also by Jensen and Falkenberg (1942), is equally incorrect. Presence of a hydrogen on the nitrogen adjacent to the SO_2 group is not necessary for salt formation because the effect of the sulphonyl group is transmitted to the other " vinylogous " nitrogen, e.g. suphathiazole dissolves readily in dilute sodium hydroxide solution yet its hydrogen is on the nitrogen more distant from sulphur (*vide infra*).

Since mono-substituted amidines have never been isolated in two pure tautomeric forms (Shriner and Neumann 1944, p. 380), and since the evidence on the tautomerism of the sulphanilylamidines was inconclusive it was decided to reinvestigate their structure. Apart from tautomerism there was another possible explanation for the existence of two forms; they might both have the structure I, expected to be more stable than II, but might be geometrical isomers, differing in the steric arrangement around the C=N bond. Such isomerism has been observed in disulphonyl derivatives of aminothiazoles (cf. Part III, to follow).*

The two acetylsulphanilylbenzamidines are not dimorphous forms of the same substance. Their spectra are different (Fig. 1). The α-form was readily converted into the β-form, but the reverse conversion could not be achieved. Both forms dissolved in pyridine but on dilution with water each solution deposited the β-form. Cold concentrated sulphuric acid dissolved both forms, but on pouring on to ice each solution gave the β-form (m.p. 207–208 °C.) of the deacetylated amidine. The α-form did not dissolve on prolonged shaking with 2·5N potassium hydroxide solution, but, surprisingly, the β-form was readily soluble in this solution and was recovered unchanged on neutralization. The reason for this behaviour is not clear ; neither the α- nor the β-form of sulphanilylbenzamidine dissolved in alkali.

---

*Barber (1943) obtained two forms of ethyl acetylsulphanilylbenzimino-ether (V), a compound in which tautomerism is not possible. We found, however, that the two forms could be produced at will by inoculation with the respective crystals; and the absorption spectra in the ultraviolet were identical. Hence, as Barber suspected, this is a case of dimorphism.

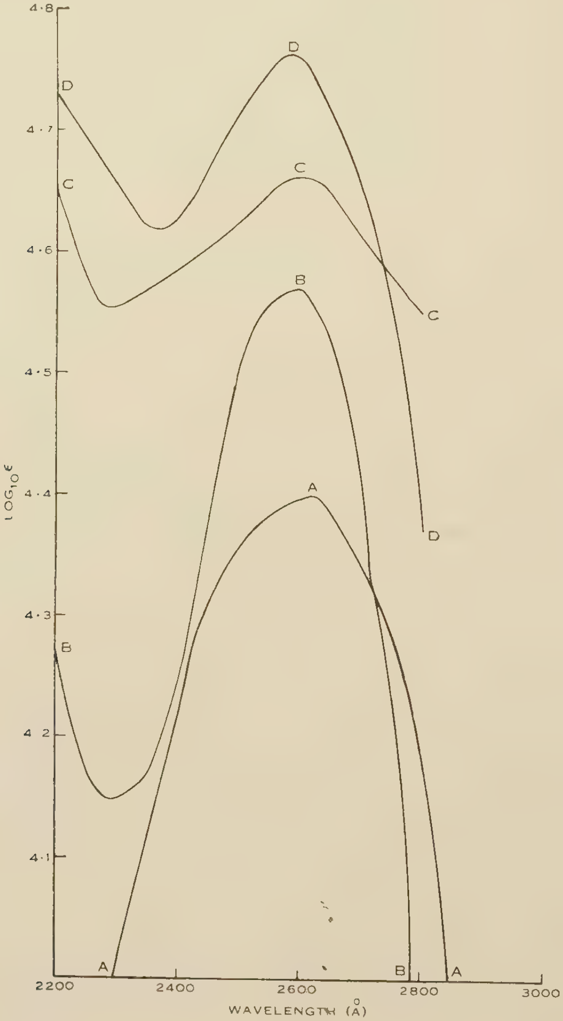
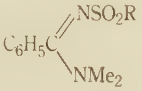


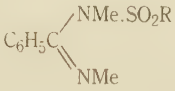
Fig. 1.—Ultraviolet absorption spectra of acetylsulphanilylbenzamidine (Curve *A*, α-form ; Curve *B*, β-form) and sulphanilylbenzamidine (Curve *C*, α-form ; Curve *D*, β-form) in ethanol.

The two forms of sulphanilylbenzamidine also have different spectra (Fig. 1). All these spectra are dominated by the strong " sulphanilamide peak " at about 2600 Å which masks any information concerning the amidine part of the molecule. To suppress the sulphanilamide peak, the absorption spectra of the sulphanilylbenzamidines were taken in acid solution, but both forms gave the same spectrum indicating that the α-form had been changed to β- by hydrogen ions.

To obtain more information from the spectra, analogues with known structure were prepared, i.e. the N,N- (VI) and the N,N′-*dimethyl derivatives* (VII) of acetylsulphanilylbenzamidine. The former was obtained, from V and dimethylamine, in two forms ; α, m.p. 174–175 °C. and β, m.p. 193·5–194·5 °C. These two forms, of the same composition, had similar but not identical absorption spectra (Fig. 2) and could not be interconverted either by 3 hours' boiling in pyridine solution, nor by 3 hours' heating at 209 °C.



(VI)



(VII)

Neither form was identical with the *N*,*N*′-dimethyl derivative described below. They must therefore be the *cis-trans* isomers of VI, showing not only that such

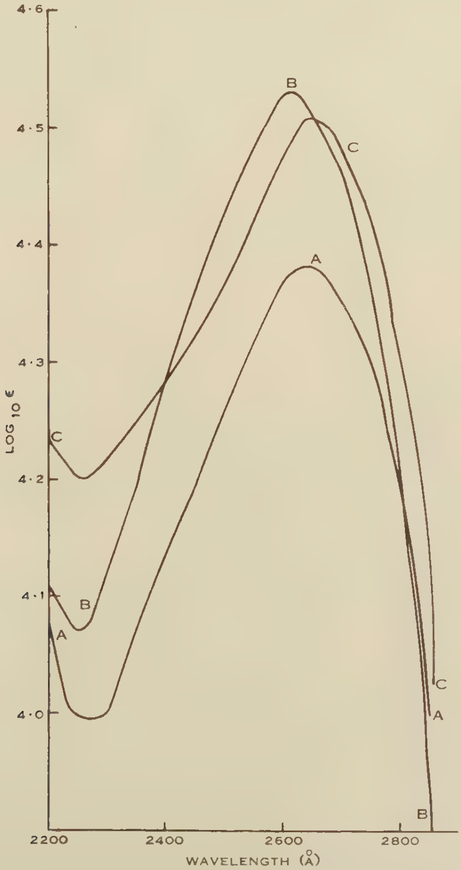


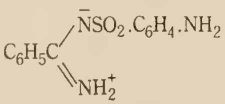
Fig. 2.—Ultraviolet absorption spectra of *N*,*N*′-dimethylacetylsulphanilylbenzamidine (Curve *A*) and *N*,*N*-dimethylacetylsulphanilylbenzamidine (Curve *B*, α-form ; Curve *C*, β-form) in ethanol.

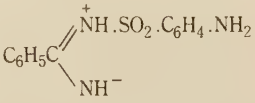
geometrical isomers can be isolated but also that the difference in their spectra is less than in the case of the two forms of the unmethylated compounds.

The *N*,*N'*-isomer (VII) was prepared from *N*,*N'*-dimethylbenzamidine* and acetylsulphanilyl chloride by a modification of the method of Northey, Pierce, and Kertesz (1942).

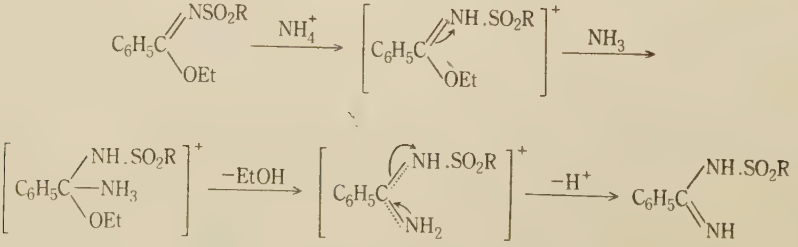
Comparison of the absorption spectra of VI and VII (Fig. 2) shows the influence of the position of the double bond. The isomer with the double bond nearer to the sulphonamide group (VI) has a stronger absorption peak (log ε about 0·14 greater) which is at a slightly lower wavelength. The spectra of the two forms of acetylsulphanilylbenzamidine and of sulphanilylbenzamidine show the same relationship: the maximum values of log ε of the β-forms are 0·17 and 0·10 higher, and at a somewhat lower wavelength. We conclude therefore that *the two forms are tautomers*, as Barber suspected, but the higher melting forms have structure I and the lower II.

The higher absorption peak of the β-form can probably be explained on the basis of the theory that high intensity of absorption is associated with a large charge separation in resonating compounds (Pauling 1943). The sulphanilamide peak is due to the resonance which makes the amino group positive and the sulphonamide group negative. In the β-forms contribution of I*a* to the resonance will increase the negative charge on the sulphonamide group, resulting in a higher dipole moment, and stronger absorption. This is also apparent in the uniformly higher melting points of the substances having the double bond nearer to the sulphonyl group. In the α-forms the direction of the electron distribution due to the amidine resonance is opposed to the electron-attraction of the sulphonyl group, and therefore contribution of II*a* will be a minor one.


(Ia)


(IIa)

Formation of the unstable tautomer from the imino-ether (V) and ammonia can be explained on the basis of the mechanism suggested by Barber, Wilkinson, and Edwards (1947).


(VIII)

---

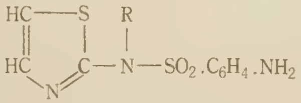
* *N*,*N'*-Dimethylbenzamidine has been described only once, by Pyman (1923), who obtained it in 15 per cent. yield in the methylation of benzamidine; its separation from a mixture of five products was very tedious. A suitable method of preparation was found in the reaction of *N*-methylbenzimino-chloride (Pechmann 1895) with cold aqueous methylamine.

The mesomeric amidinium ion (VIII) will, because of the electron-attraction of the SO_2 group, have a low electron density on the amino group as indicated by the curved arrows. A proton will be lost from this group giving the α-form of the amidine. This process is similar to the initial formation of the less stable tautomer, the enol, from the anion of a keto-enol system, or of the *aci*-form from the anion of a nitro compound. All these cases are examples of kinetic, as opposed to thermodynamic, control of reaction products (cf. Catchpole, Hughes, and Ingold 1948).

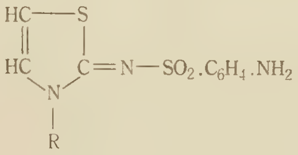
Occurrence of amidines in tautomeric forms capable of isolation can be explained as follows. Amidines usually tautomerize at a fast rate presumably either by proton addition and elimination, or by association (Shriner and Neumann 1944, p. 380), both of which are due to the strongly basic nature of the amidines. The powerful sulphonyl group, however, has reduced the basicity, and thereby the rate of interconversion, to such an extent that, in the absence of basic or acid catalysts, the labile forms can be isolated.

III. SULPHATHIAZOLE AND SULPHAPYRIDINE

In these compounds tautomerism is also possible, e.g. in sulphathiazole between IX and X.



(IX, R=H)
(XI, R=Me)



(X, R=H)
(XII, R=Me)

That they are really tautomeric is indicated by the fact that reaction with diazomethane gives both possible methyl derivatives, e.g. from sulphathiazole (XI and XII); according to Arndt (1936) this is evidence for tautomerism.

It has been claimed (Bednyagina and Postovskii 1946) that the ratio of the methylated isomers represents the ratio of the tautomers in the parent substance, a claim totally unjustified except in the unlikely case that both tautomers were methylated at the same rate, and that this rate was very great compared with that of the tautomeric rearrangement.

Further information about the tautomerism can be gained from comparison of the absorption spectra of the parent compounds with those of the two methylated isomers. This has been done by Shepherd, Bratton, and Blanchard (1942) for sulphapyridine and sulphathiazole. We have redetermined these spectra which are shown, in case of the thiazoles, by curves *A* and *B* in Figure 3. The spectra of sulphathiazole and XII were so nearly identical that separate curves could not be drawn. The curve for XI shows a definite inflection at about 2620 Å which is not evident in Shepherd's work; it persisted after repeated purification and is therefore not due to an impurity. From these spectra Shepherd, Bratton, and Blanchard (1942) concluded that sulphathiazole is nearly completely in the " imino " form (X). They assigned the 2850–2900 Å band to the sulphanilamide half of the molecule since this band occurs in all the spectra, and the peak at 2600 Å to the 2-thiazolone-imine structure, since it was not evident in the spectrum of XI and since 3-methyl-2-thiazolone-imine was found to absorb at about the same wavelength.

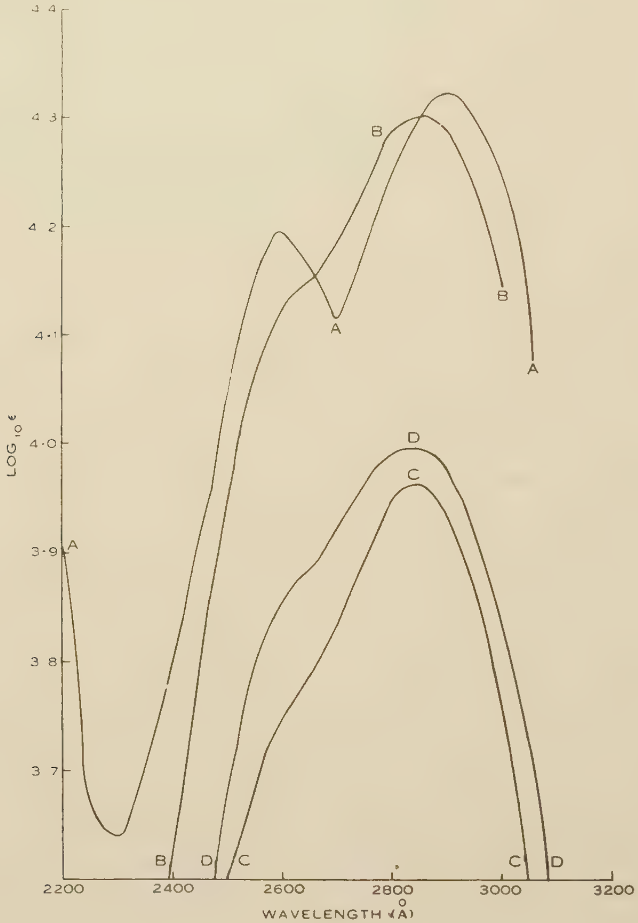
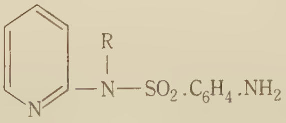


Fig. 3.—Ultraviolet absorption spectra of sulphathiazole and of its methyl-homologues. Curve *A*, 3-Methyl and parent compounds ; Curve *B*, N^1-methyl compound in ethanol. Curve *C*, 3-Methyl and parent compounds ; Curve *D*, N^1-Methyl compound in 0·05N 50 per cent. ethanolic hydrochloric acid.

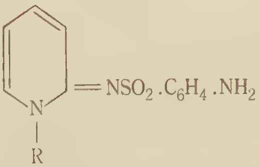
These conclusions were strongly attacked by Vandenbelt and Doub (1944), who pointed out that the absorption maximum of 2-aminothiazole is also near to 2600 Å ; that it was unlikely that the position of this maximum would be unaffected by sulphanilyl substitution, and that in other related compounds the peak at 2600 Å was identified with the sulphanilamide structure. In a thorough study of the absorption spectra of several sulpha drugs, Vandenbelt and Doub conclusively proved that the 2600 Å peak was due to the sulphanilamide portion of the molecule: it was considerably reduced by taking the spectrum in acid solution. This is a well-known characteristic of aromatic amines (Kumler and Strait 1943). With the interpretation of their spectra thus invalidated, Shepherd's correct conclusion, *that sulphathiazole is mainly X*, was generally disregarded. It was also left unexplained why the sulphanilamide peak should have disappeared from the spectrum of XI.

We have determined the spectra of sulphathiazole and its two methyl derivatives (Shepherd, Bratton, and Blanchard 1942) in acid solution (Fig. 3). Again, the curves for sulphathiazole and XII were indistinguishable, indicating that in acid solution also sulphathiazole is mainly in form X. Disappearance of the 2600 Å band proves that this is the "sulphanilamide peak"; the band at 2840 Å is due to the thiazole moiety and indicates that the aminothiazole and the thiazolone-imine structures have substantially the same absorption characteristics.* Disappearance of the sulphanilamide peak from curve *B* can be explained by the results of the previous section: if the conclusions reached on the spectra of sulphanilylamidines are valid in the case of sulphathiazole (which is a cyclic amidine), then log ε at 2600 Å in curve *B* should be about 0·14 less than in curve *A* and at a slightly longer wavelength. As is clearly visible in Figure 3, such a change in the maximum would merge it into the side of the larger absorption band.

The case of sulphapyridine is somewhat more complicated. Shepherd, Bratton, and Blanchard (1942), for reasons not explained, carried out the comparison of the absorption spectra with those of the two methyl derivatives not on sulphapyridine itself, but on the corresponding acetyl compounds. They found that all three compounds had a peak at about 2650 Å which they assigned to the sulphanilamide structure. A second band occurred at 3215 Å in the acetyl derivative of XVI but not in the acetyl derivative of XV and was therefore regarded as characteristic for the pyridone-imine structure. Acetylsulphapyridine was found to show both maxima, but the absorption at the higher wavelength was weaker than in XVI. From this evidence it was estimated that acetylsulphapyridine contains about 60 per cent. of the "imino" structure (XIV).



(XIII, R=H)
(XV, R=Me)



(XIV, R=H)
(XVI, R=Me)

We have determined the spectra of sulphapyridine itself and its two methyl derivatives in neutral (Fig. 4) and acid alcoholic solution (Fig. 5). The curves

* The decrease of this peak in acid is obviously due to the reduction of the 2600Å band which contributed considerably to the absorption even at 2840 Å (Vandenbelt and Doub 1944).

differ somewhat from those of the acetyl derivatives. The nuclear-methylated isomer (XVI) shows three peaks of which the one at 2650 Å disappears in acid solution and is therefore due to the sulphanilamide moiety of the molecule; the other two, at 2500 and 3250 Å, are characteristic for the pyridone-imine structure. The N^1-methylated isomer (XV) shows only one broad band which is a combination of the sulphanilamide peak and another peak at about 2750 Å, characteristic for the aminopyridine structure. Again, like in the other compounds studied, the "sulphanilamide peak" is higher in the isomer with the double bond nearer to the sulphonamide group, and at a lower wavelength.

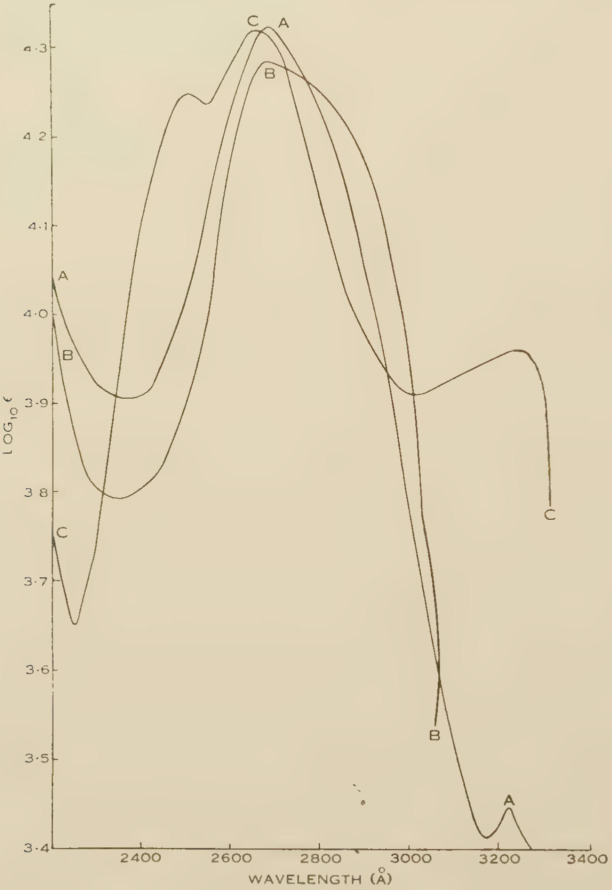


Fig. 4.—Ultraviolet absorption spectra of sulphapyridine and of its methyl-homologues in ethanol. Curve *A*, Parent compound; Curve *B*, N^1-Methyl compound; Curve *C*, 1-Methyl compound.

The curves for sulphapyridine show that it is a *mixture of the two tautomeric forms*; and assuming that the absorption coefficients of these forms are approximately the same as those of the corresponding methyl derivatives, it can be calculated that about 30 per cent. of the mixture is in the " imino " form (XIV). Similar calculation on the spectra in acid solution indicates 70 per cent. XIV.*

Thus, while in sulphathiazole the electron-attraction of the sulphonyl group more than compensates for the reduction of resonance in the " imino " form, in sulphapyridine, the resonance energy of pyridine being much higher than of thiazole, the two factors nearly balance each other and both tautomers will be present in considerable amounts. Electron-attracting groups in the *p*-position will increase the electron-attraction of the sulphonyl group ; hence acetylation of the *p*-amino group increases the proportion of XIV, and addition of a proton to the amino group will increase it further.

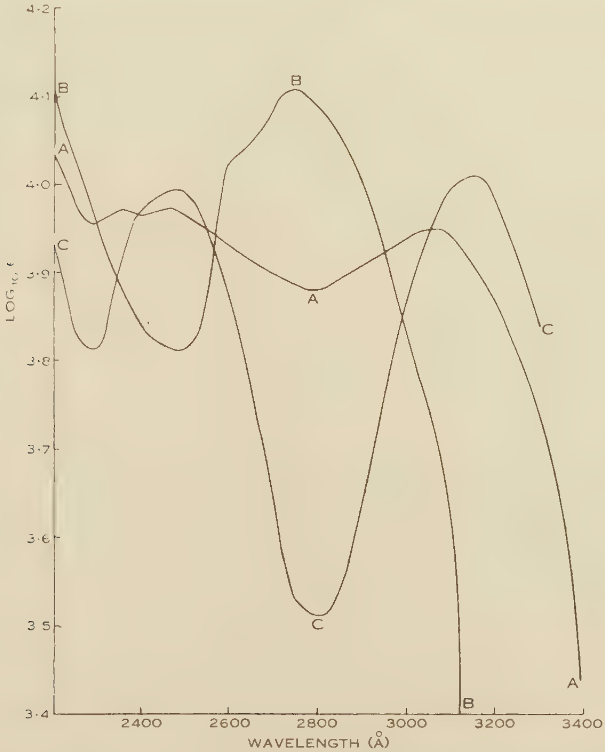


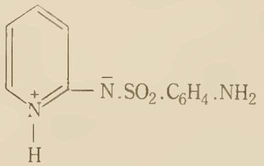
Fig. 5.—Ultraviolet absorption spectra of sulphapyridine and of its methyl-homologues in 0·5N 50 per cent. ethanolic hydrochloric acid. Curve *A*, Parent compound ; Curve *B*, N^1-Methyl compound ; Curve *C*, 1-Methyl compound.

Vandenbelt and Doub's (1944) curve for sulphapyridine, taken in aqueous solution, shows a higher peak at 3200 Å than our spectrum, taken in ethanol. It is not possible to calculate the percentage of " imino " form (XIV) from their curve because the spectrum of XVI in aqueous solution is not known ; but it is clear that it is much higher in aqueous than in alcoholic solution. This is in accord with the experience in the keto-enol field that the percentage of the more polar form (the ketone) increases with increasing polarity of the solvent.

---

* The former figure is the ratio of the ε values at 3230 Å, the latter was averaged from the values at 2760 and 3060 Å.

The above conclusions, drawn from the study of spectra, are valid only for the very dilute ethanolic solutions in which the spectra were taken. They give no information on the structure of crystalline sulphapyridine or -thiazole. Comparison of the physical properties, however, makes it very likely that both solid compounds are in the " imino " form (X and XIV) respectively. Both parent compounds and their ring-methylated derivatives have high melting points (above 190 °C.) and are insoluble in ether, while the N^1-methylated compounds melt much lower (87 and 110 °C.) and are soluble in ether. These differences are, of course, due to resonance: structures of the type XVII will contribute considerably to the electron-distribution in the first-mentioned molecules because they contain a fully aromatic ring and satisfy the electron-attraction of the sulphonyl group. The higher negative charge on the sulphonyl group causes greater polarity which is responsible for the high melting point and low solubility in ether. Resonance in the " amino " forms is much smaller because charge separation here weakens the aromatic resonance. It is interesting that the ring-methylated derivatives are much more active therapeutically than the N^1-methylated isomers (Shepherd, Bratton, and Blanchard 1942).



(XVII)

IV. Experimental

Melting-points are corrected. Microanalyses by Miss B. Naylor.

(a) *Ethyl Acetylsulphanilylbenzimino-ether* (*V*)

This compound was prepared by the method of Barber (1943); adherence to the conditions given here greatly improved the yield. Freshly prepared benzimino-ether hydrochloride (7·9 g.) was suspended in anhydrous pyridine (20 ml.) and to it added in one portion *p*-acetamidobenzene-sulphonyl chloride (9·8 g.) containing less than 0·15% water. The flask was immersed in boiling water until the reaction mixture reached 70 °C. and was kept at that temperature for 2 minutes with stirring. It was cooled rapidly and poured into ice and water (500 ml.). After 30 minutes the precipitated solid was collected and recrystallized from cold benzene (40 ml.) with light petroleum (b.p. 60–90 °C., 40 ml.) to give colourless plates (3·6 g., 25%), m.p. 99–102 °C. Recrystallization from aqueous ethanol gave m.p. 100–102 °C.

It was not found possible to condense the imino-ether itself with this sulphonyl chloride in acetone using one excess mole of the imino-ether as a condensing agent, a general method described by Barber (1943), and pyridine did not promote the reaction.

(b) N,N^1-*Dimethyl-acetylsulphanilylbenzamidine* (*VI*)

The above compound (1·5 g.) was dissolved in anhydrous ethanol (20 ml.) and dry dimethylamine (*c.* 1 g.) was passed in with occasional shaking during about 15 minutes. The reaction mixture warmed slightly and the amine was readily absorbed. The solution was concentrated to 5 ml. and allowed to crystallize. The *product* was recrystallized once from ethanol to give colourless plates, m.p. 194–195 °C. (0·60 g., 40%); this was called the β-form.

Found : N, 12·3%.

Calculated for $C_{17}H_{19}O_3N_3S$: N, 12·2%.

When the preparation was repeated under substantially the same conditions the recrystallized *product*, obtained in 60% yield, had m.p. 173·5–174·5 °C. This was called the α-form.

Found : N, 12·2%.

Various attempts, using carefully controlled conditions, to repeat these results, led to the recovery of one form or the other in 40–60% yield, and from the mother liquors of one there was obtained, on concentration, a small percentage of the other. The total yield of products was usually about 70%.

(*c*) N,N¹-*Dimethylbenzamidine Hydrochloride*

Methylamine hydrochloride (15·0 g.) was dissolved in a solution of pure sodium hydroxide (8·7 g.) in water (60 ml.) and cooled to below 10 °C. To this solution was added *N*-methylbenzimido-chloride (14·0 g.) with cooling in salt and ice, the addition occupying 6 minutes and the temperature being kept below 10 °C. The mixture was then removed from the cooling bath and after 10 minutes was made alkaline with 10% sodium hydroxide. An oil separated and was extracted with chloroform (3 portions of 50 ml.). The extract was dried with sodium sulphate and the chloroform removed. The oily residue was dissolved in dry ether (80 ml.) and dry hydrogen chloride passed in, with stirring, for 30 minutes, moisture being excluded. The hydrochloride was collected and recrystallized as deliquescent prisms from anhydrous ethanol (60 ml.) by adding ether (20 ml.) and cooling in salt and ice to give a microcrystalline product (7·2 g., 40%), m.p. 255–256 °C. The picrate was prepared, m.p. 171–172 °C., and also the free amidine, m.p. 80–81 °C. Pyman (1923) records the same m.p.'s for all three compounds.

(*d*) N,N¹-*Dimethyl-acetylsulphanilbenzamidine* (*VII*)

N,*N*′-Dimethylbenzamidine hydrochloride (1·23 g.) was suspended in acetone (13 ml.) and to it added a solution of sodium hydroxide (0·63 g.) in water (1 ml.). To this mixture was added at room temperature during 3 minutes *p*-acetylsulphanilyl-chloride (1·67 g.) with shaking. The mixture was then shaken for 1 hour at room temperature and ice-cold water (35 ml.) added whereupon an oil separated. After having kept the mixture overnight in the refrigerator, the aqueous layer was decanted, the oil washed with ice-cold water, and stirred until it solidified. It was recrystallized from 50% ethanol (20 ml.) to give colourless *plates* (1·40 g., 68%), m.p. 154–155 °C., unchanged by recrystallization from aqueous acetone.

Found : N, 12·1%.

Calculated for $C_{17}H_{19}O_3N_3S$: N, 12·2%.

(*e*) N¹-*Methyl-2-sulphanilamidopyridine* (*XV*)

This compound could not be obtained by treatment of sulphapyridine with diazomethane according to Shepherd, Bratton, and Blanchard (1942). It was prepared by methylation of acetylsulphapyridine, followed by hydrolysis. To an ethereal solution of diazomethane, prepared from nitrosomethylurea (2·5 g.), acetylsulphapyridine (3·7 g.), and methanol (1 ml.) were added. After shaking for 4 hours the solution had lost its colour. The solid product was filtered off and extracted twice with ether (50 ml.) containing 3% ethanol. The filtrate and the extracts were combined and evaporated ; addition of methanol caused separation of crystals (1·0 g.). These were hydrolysed by 2 hours' boiling with a solution of sodium hydroxide (0·3 g.) in ethanol (8 ml.). Dilution with water precipitated an oil which rapidly crystallized ; recrystallization from aqueous methanol gave long needles (0·5 g.), m.p. 83 °C. Shepherd, Bratton, and Blanchard (1942) report m.p. 86·5–87 °C.

(*f*) *1-Methyl-2-sulphanilimido-1,2-dihydropyridine* (*XVI*)

Sulphapyridine (2·5 g.) was dissolved in the minimum volume of 10% sodium hydroxide solution, and dimethyl sulphate (1·2 g.) added at room temperature dropwise with stirring, the mixture being kept just alkaline to litmus. An oil separated and was washed with water. It solidified in the refrigerator and was recrystallized twice from 6N acetic acid to give colourless needles (1·8 g., 70%), m.p. 231–232 °C. Shepherd, Bratton, and Blanchard (1942) report m.p. 232–233 °C.

(g) *Ultraviolet Spectra*

The spectra were determined on a Beckmann quartz spectrophotometer with a hydrogen-discharge tube as a light source, using a 1 cm. cell. Concentrations were 1 or 2 mg./100 ml.

Details are given below for the spectra not shown in graphs :

Ethyl acetylsulphanilylbenzimino-ether in ethanol : maximum at 2610 Å, log ε 4·46, minimum at 2250 Å, log ε 4·01.

Acetylsulphanilylbenzamidine, β-form, in 50% alcoholic N sodium hydroxide : log ε 4·50 at 2570 Å.

Sulphanilylbenzamidine in 50% alcoholic 2N hydrochloric acid : broad maximum between 2360 and 2600 Å with log ε 4·81 at 2400 Å.

Ethyl benzimino-ether in ethanol : broad maximum with log ε 4·07 at 2280 Å.

N,*N′*-Dimethylbenzamidine in ethanol : broad maximum with log ε 4·05 at 2230 Å.

The absorption of benzimino-ethers and benzamidines has apparently not been recorded before. Ethyl acetimino-ether has been stated to have no absorption above 2200 Å (Glickman and Cope 1945).

V. ACKNOWLEDGMENTS

The authors are indebted to Messrs. Andrew's Laboratories and to Messrs. May and Baker (Australia) Pty. Ltd. for gifts of chemicals.

VI. REFERENCES

ARNDT, F. (1936).—*Rev. Fac. Sci. Univ. Istambul* **1** : 1.

ARNDT, F., and SCHOLZ, H. (1934).—*Liebigs Ann.* **510** : 62.

BACKER, H. J., and MOED, H. D. (1947).—*Rec. Trav. Chim. Pays-Bas* **66** : 335.

BARBER, H. J. (1943).—*J. Chem. Soc.* **1943** : 101.

BARBER, H. J., WILKINSON, J. H., and EDWARDS, W. G. H. (1947).—*J. Soc. Chem. Ind.* **66** : 411.

BEDNYAGINA, N. P., and POSTOVSKII, I. YA. (1946).—*J. Gen. Chem. U.S.S.R.* **16** : 1941.

BRANCH, G. E. K., and CALVIN, M. (1941).—"The Theory of Organic Chemistry." p. 289. (Prentice-Hall, Inc. : New York.)

CATCHPOLE, A. G., HUGHES, E. D., and INGOLD, C. K. (1948).—*J. Chem. Soc.* **1948** : 8.

GLICKMAN, S. A., and COPE, A. C. (1945).—*J. Amer. Chem. Soc.* **67** : 1017.

JENSEN, K. A., and FALKENBERG, P. (1942).—*Dansk. Tids. Farm.* **16** : 145.

KUMLER, W. D., and DANIELS, T. C. (1943).—*J. Amer. Chem. Soc.* **65** : 2190.

KUMLER, W. D., and STRAIT, L. A. (1943).—*J. Amer. Chem. Soc.* **65** : 2349.

MARKWALD, W. (1895).—*Liebigs Ann.* **286** : 343.

NORTHEY, E. H. (1948).—"The Sulfonamides and Allied Compounds." p. 27. (Reinhold Publishing Corp. : New York.)

NORTHEY, E. H., PIERCE, H. E., and KERTESZ, D. J. (1942).—*J. Amer. Chem. Soc.* **64** : 2763.

PAULING, L. (1943).—"Gilman's Organic Chemistry." 2nd Ed. p. 1982. (John Wiley & Sons, Inc. : New York.)

PECHMANN, H. v. (1895).—*Ber. dtsch. chem. Ges.* **28** : 2367.

PYMAN, F. L. (1923).—*J. Chem. Soc.* **1923** : 3365.

SHEPHERD, R. C., BRATTON, A. C., and BLANCHARD, K. C. (1942).—*J. Amer. Chem. Soc.* **64** : 2532.

SHRINER, R. L., and NEUMANN, F. W. (1944).—*Chem. Rev.* **35** : 384.

VANDENBELT, J. M., and DOUB, L. (1944).—*J. Amer. Chem. Soc.* **66** : 1633.

WHEELER, H. L., JOHNSON, T. B., and MCFARLAND, D. F. (1903).—*J. Amer. Chem. Soc.* **25** : 790.

THE CHEMISTRY OF WESTERN AUSTRALIAN PLANTS

IV. DUBOISIA HOPWOODII

By W. BOTTOMLEY* and D. E. WHITE†

[*Manuscript received July* 27, 1950]

Summary

Examination of a large number of samples of *Duboisia hopwoodii* has confirmed the findings of Bottomley, Nottle, and White (1945). The methods used for the estimation of nicotine and nornicotine are described and the results summarized. These could not be correlated with locality, season, or soil type, and examination of fresh leaves showed that the alkaloid content is considerably reduced on drying.

Samples of nicotine and nornicotine have been prepared and ursolic acid and a quercetin glycoside have been isolated from *D. hopwoodii* leaves.

I. INTRODUCTION

The genus *Duboisia* (family Solanaceae) includes only three species, all of which are indigenous to the Australian mainland. Two of these, *D. myoporoides* R.Br. and *D. leichhardtii* F. Muell., are known to contain only tropane alkaloids, principally hyoscyamine and hyoscine (Henry 1949). The third species, *D. hopwoodii* (F. Muell.) F. Muell., a tall shrub with glabrous lanceolate leaves and small bell-shaped flowers, is found in the more arid parts of Australia and for many years has been known to be poisonous to stock. Its use by the aborigines, under the name " pituri ", to poison water holes in order to facilitate emu hunting and to chew for its narcotic effects has also been reported (Bancroft 1872, 1877 ; Liversidge 1880 ; Hicks and Le Messurier 1935).

Crude alkaloidal preparations similar to Bancroft's (1872, 1877) " piturine " were examined by Ringer and Murrell (1878–1879), Fraser (1878–1879), Langley and Dickinson (1890), Petit (1879), and Rothera (1910), the last named obtaining strong evidence for the identity of " piturine " and nicotine.

Subsequently, Hicks and Le Messurier (1935), Späth, Hicks, and Zajic (1935), Hicks (1936), Bowen (1944), and Hicks and Sinclair (1947) found nornicotine to be the major alkaloid, but Grantham (1943) identified nicotine as its picrate in several samples and Trautner and Neufeld (1946) stated that nicotine was dominant in samples from Queensland and Western Australia.

Our examination of Western Australian specimens of *D. hopwoodii* has demonstrated that nicotine and nornicotine are usually both present (Bottomley, Nottle, and White 1945). This finding has now been confirmed and extended by the examination of 67 specimens. Material from only four of the 34 sites where specimens were collected showed a complete absence of nicotine. The

---

* Present address : Division of Plant Industry, C.S.I.R.O., Canberra, A.C.T.

† Organic Chemistry Department, University of Western Australia, Nedlands, W.A.

nicotine contents found ranged from 0 to 5·3 per cent. (mean 1·28 per cent.) and the nornicotine contents were between 0·1 and 4·1 per cent. (mean 1·17 per cent.). Although nicotine was absent from four samples nornicotine was found in all the samples examined and only six contained less than 0·4 per cent.

The nicotine and nornicotine were estimated by the method outlined previously (Bottomley, Nottle, and White 1945) and now described in Section II. Although probably less accurate than the procedures of Bowen and Barthel (1943), Bowen (1947), and Larson and Haag (1944), this method was simple, rapid, and sufficiently accurate for our purpose.

Although samples were examined from widely separated localities and a variety of different soils at intervals between January 1945 and April 1946 no correlation between locality, season or soil type, and alkaloid content could be detected and it was eventually found that the results could not be used for this purpose, as the alkaloid content of the leaves was subject to large and unpredictable losses on drying. This was established by the examination of the samples listed in Table 1 which were halved, half being examined fresh, within an hour of collection, and the other half after drying to constant weight at 35 °C. This required laboratory facilities close to the site of the plants and hence it was only possible to examine four samples which included none with a high alkaloid content as estimated on the dry material. Thus it is not known whether the total alkaloid content ever rises above 6 per cent. or whether this is a maximum value.

TABLE 1

COMPARISON OF THE ALKALOID CONTENT OF THE LEAVES OF D. HOPWOODII BEFORE AND AFTER DRYING

Sample No.	Nicotine		Nornicotine	
	Fresh (%)	Dried (%)	Fresh (%)	Dried (%)
215	4·8	1·0	0·9	0·3
282	3·4	1·1	0·5	0·05
283	3·0	2·1	0·3	0·1
284	4·7	0·9	1·3	0·6

Extraction of a mixed bulk sample of *D. hopwoodii* leaves gave pure (−)-nicotine and a higher boiling fraction containing nornicotine. Pure nornicotine was obtained from a nicotine-free plant specimen and like the product obtained by Späth, Hicks, and Zajic (1935, 1936) was a mixture of (+)- and (±)-forms.

During the concentration of the alcoholic extract of the alkaloids a solid separated and, on purification, was identified as ursolic acid by comparison of the acid and its methyl ester with authentic specimens. Trautner and Neufeld

(1947) also reported this triterpene acid in *D. hopwoodii* as well as the other two *Duboisia* species but did not give a clear statement of the yield obtained.

On extraction of the concentrated acidified extract with chloroform, a yellow solid separated and was identified as a quercetin glycoside which on hydrolysis gave quercetin, characterized as its penta-acetyl derivative and compared with an authentic sample.

II. Experimental

(*a*) *Estimation of Nicotine*

To the dried, ground leaves of *D. hopwoodii* (20 g.) were added calcium hydroxide (5 g.) and water (50 ml.) and the mass distilled in a current of steam, the distillate being collected under 2N hydrochloric acid. Care was taken to avoid superheating of the distilling flask, and the distillation was continued until a test portion of the distillate gave no opalescence with a saturated picric acid solution. Excess saturated picric acid solution was then added to the distillate and allowed to stand overnight. The precipitate of nicotine dipicrate was filtered, washed with a little water, dried at 105 °C., and weighed.

$$\text{Percentage nicotine} = \text{mass of dipicrate} \times 1\cdot306.$$

The picrates obtained all melted at 217 °C. or higher and on one crystallization from water formed yellow needles, m.p. 222–224 °C. (decomp.) undepressed by an authentic sample of nicotine dipicrate, m.p. 223–224 °C.

(*b*) *Estimation of Nornicotine*

Salt (20 g.) and sodium hydroxide (5 g.) were added to the distillation residue after the removal of the nicotine, and the distillation continued until a test portion of the distillate gave no turbidity with silicotungstic acid solution. An aliquot (usually about 120 ml.) of the distillate was acidified with 10N hydrochloric acid (4 ml.), excess 12% silicotungstic acid solution carefully added, and allowed to stand for 1 hour. The precipitate of nornicotine silicotungstate was filtered, washed free from chloride, ignited in a crucible at 600–700 °C., and weighed.

$$\text{Mass nornicotine} = \text{mass residue} \times 0\cdot1042 \text{ (Bowen and Barthel 1943).}$$

(*c*) *Extraction of Nicotine and Nornicotine*

Dried, ground leaves of *D. hopwoodii* (5·3 kg.), collected from various localities, were percolated continuously with alcohol for 200 hours. The extract was acidified with acetic acid, reduced to small bulk by evaporation under reduced pressure, and filtered, when a green amorphous precipitate (*U*) was obtained. The filtrate was acidified to congo red with hydrochloric acid and exhausted with chloroform. During this extraction a yellow solid (*Q*) separated from the aqueous phase and was removed by filtration. The aqueous solution was then carefully made alkaline with sodium hydroxide, again extracted with chloroform, the chloroform evaporated, and the residue distilled under reduced pressure through a column 30 by 2 cm. packed with glass helices.

Fraction I, b.p. 76·5 °C./3 mm. (39 g.; 0·74%),
Fraction II, b.p. 76·5–96 °C./3 mm. (7 g.),
Fraction III, b.p. 96–96·5 °C./3 mm. (15 g.; 0·28%).

(i) (—)-*Nicotine*.—Fraction I was redistilled.

Found: C, 74·1; H, 8·7; N, 17·6%; M, 161.

Calculated for $C_{10}H_{14}N_2$: C, 74·0; H, 8·7; N, 17·3%; M, 162·2.

$[\alpha]_D^{20}$ —164·8°.

The dipicrate was prepared by treating an aqueous solution of (—)-nicotine with an excess of a saturated aqueous picric acid solution. The precipitate of (—)-nicotine dipicrate formed yellow needles from aqueous alcohol, m.p. 222–223 °C. alone or mixed with authentic (—)-nicotine dipicrate.

(ii) *Nornicotine.*—Fraction III still contained nicotine and exhibited a specific rotation $[\alpha]_D^{20}$ —6°. As it was found difficult to purify by distillation, a pure sample of nornicotine was obtained from a sample of *D. hopwoodii* which had been shown to contain no nicotine. The extraction was carried out as described above, and the oil, on distillation, yielded only one fraction, b.p. 104·5–106 °C./4 mm.

Found: C, 73·3; H, 8·2%.

Calculated for $C_9H_{12}N_2$: C, 72·9; H, 8·2%.

$[\alpha]_D^{20}$ +30·5°.

The dipicrate, prepared by adding excess aqueous picric acid solution to an aqueous solution of nornicotine crystallized from aqueous alcohol, m.p. 190–191 °C., while the dipicrolonate, which was obtained in a similar way using saturated aqueous picrolonic acid solution, and crystallizing from alcohol, melted at 251–252 °C.

(*d*) *Ursolic Acid*

The precipitate (*U*) from the extraction of the alkaloids, was triturated with cold 5% potassium hydroxide in 50% alcohol, filtered, and the filtrate acidified with a mixture of hydrochloric acid and alcohol (1 : 1). The precipitate (38 g.; 0·7%) was filtered and recrystallized several times from alcohol when it yielded colourless needles, m.p. 289–291 °C. (corr.) undepressed by admixture with authentic ursolic acid.

(*e*) *Methyl Ursolate*

Ursolic acid (0·5 g.) was treated with potassium hydroxide (1 g.) and methyl sulphate (1·5 ml.) in aqueous alcohol (1 : 1; 15 ml.), for 2 hours, the suspension of methyl ursolate filtered, washed with aqueous alcohol, recrystallized from 75% alcohol, and dried in a vacuum at 110 °C., m.p. 171–171·5 °C. alone or mixed with authentic material.

(*f*) *Quercetin Glycoside*

The precipitate (*Q*) (21 g.; 0·4%) formed an amorphous yellow powder from water, m.p. 185–200 °C.

(*g*) *Quercetin*

The glycoside (1 g.) in 3·5N sulphuric acid (30 ml.) was boiled under reflux for 1 hour, the solution cooled, and filtered. The quercetin was crystallized from aqueous alcohol (3 : 1) and formed minute yellow crystals, m.p. 309–312 °C. with decomposition, undepressed by an authentic specimen.

(*h*) *Penta-Acetylquercetin*

Quercetin (0·2 g.) was acetylated with acetic anhydride (5 ml.) and pyridine (3 drops) under reflux for 40 minutes and the solution poured into water. The resulting solid formed colourless needles from alcohol (50 ml.), m.p. 193·5–194 °C., alone or mixed with authentic penta-acetylquercetin.

III. ACKNOWLEDGMENTS

Thanks are due to Mr. C. A. Gardner, Government Botanist, W.A., and Mr. R. D. Royce for the collection and identification of the material, to the Drug Panel of the Western Australian Department of Industrial Development for a maintenance grant, to the State (W.A.) Alunite Industry for laboratory facilities at Chandler for the work on fresh leaves, to the Commonwealth Research Grant to Australian Universities for equipment, to Mr. G. K. Hughes for an authentic specimen of the quercetin glycoside, rutin, and to Dr. Weiler and Dr. Strauss for microanalyses.

IV. REFERENCES

BANCROFT, J. (1872).—*Trans. Phil. Soc. Qd.* **1**: 2 pp.

BANCROFT, J. (1877).—*Trans. Phil. Soc. Qd.* **2**: 13 pp.

BOTTOMLEY, W., NOTTLE, R. A., and WHITE, D. E. (1945).—*Aust. J. Sci.* **8**: 18.

BOWEN, C. V. (1944).—*J. Econ. Ent.* **37**: 293.

BOWEN, C. V. (1947).—*J. Ass. Off. Agric. Chem. Wash.* **30**: 315.

BOWEN, C. V., and BARTHEL, W. F. (1943).—*Industr. Engng. Chem.* (Anal. Ed.) **15**: 740.

FRASER, T. R. (1878-1879).—*Proc. Roy. Soc. Edinb.* **10**: 200.

GRANTHAM, T. R. (1943).—Thesis for M.Sc. Degree, University of Melbourne.

HENRY, T. A. (1949).—" The Plant Alkaloids." 4th Ed. p. 65. (Churchill: London.)

HICKS, C. S. (1936).—*Aust. J. Exp. Biol. Med. Sci.* **14**: 39.

HICKS, C. S., and LE MESSURIER, H. (1935).—*Aust. J. Exp. Biol. Med. Sci.* **13**: 175.

HICKS, C. S., and SINCLAIR, D. A. (1947).—*Aust. J. Exp. Biol. Med. Sci.* **25**: 191.

LANGLEY, J. N., and DICKINSON, W. L. (1890).—*J. Physiol.* **11**: 265.

LARSON, P. S., and HAAG, H. B. (1944).—*Industr. Engng. Chem.* (Anal. Ed.) **16**: 86.

LIVERSIDGE, A. (1880).—*J. Roy. Soc. N.S.W.* **14**: 123.

PETIT, A. (1879).—*Pharm. J.* **9**: 819.

RINGER, S., and MURRELL, W. (1878-1879).—*J. Physiol.* **1**: 377.

ROTHERA, A. C. H. (1910).—*Biochem. J.* **5**: 193.

SPÄTH, E., HICKS, C. S., and ZAJIC, E. (1935).—*Ber. dtsch. chem. Ges.* **68**: 1388.

SPÄTH, E., HICKS, C. S., and ZAJIC, E. (1936).—*Ber. dtsch. chem. Ges.* **69**: 250.

TRAUTNER, E. M., and NEUFELD, O. E. (1946).—*Aust. Chem. Inst. J. and Proc.* **13**: 70.

TRAUTNER, E. M., and NEUFELD, O. E. (1947).—*Aust. Chem. Inst. J. and Proc.* **14**: 17.

SHORT COMMUNICATIONS

THE CHEMISTRY OF WESTERN AUSTRALIAN PLANTS

V. VELLEIN FROM VELLEIA DISCOPHORA*

By W. Bottomley† and D. E. White**

Velleia discophora F. Muell. (family Goodeniaceae), a straggling annual found from the wheat belt to the Kalgoorlie district of Western Australia and reported poisonous to stock (Watson 1943–44) has been extracted with ethanol and a compound $C_{20}H_{24}O_8.1/2H_2O$, which it is proposed to name *vellein*, has been isolated. Vellein gave a tetra-acetyl derivative and on hydrolysis with acid it formed a sugar and a compound, $C_{14}H_{14}O_3$, which did not form an acetyl derivative and was not reduced with hydrogen and a platinum catalyst but contained a coumarin ring (opened with alkali and closed again with methanolic sulphuric acid). The absorption spectrum of this compound showed a marked resemblance to that of umbelliferone (Fig. 1 and Table 1) and its melting point coincided with that of dihydroseselin (Späth *et al.* 1939).

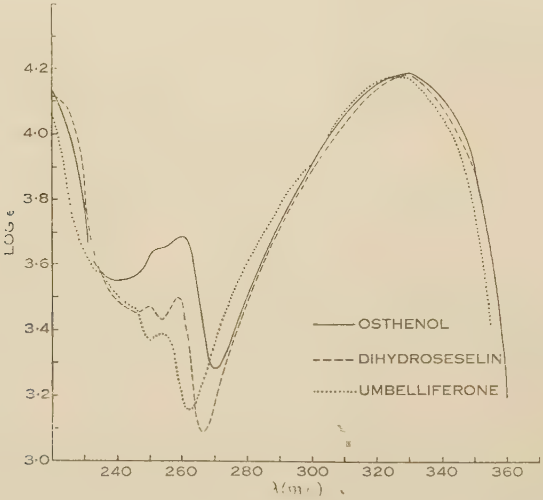


Fig. 1

* Manuscript received September 13, 1950.

† Present address : Division of Plant Industry, C.S.I.R.O., Canberra.

** Organic Chemistry Department, University of Western Australia, Nedlands, W.A.

On hydrolysis of vellein with emulsin, the phenolic coumarin osthenol (Späth and Bruck 1937) was obtained and identified by comparison of its methyl ether with an authentic specimen of osthol. Treatment of osthenol with dilute hydrochloric acid converted it to dihydroseselin (cf. Späth, Eiter, and Meinhard 1942, who used hydrobromic acid and phosphorus), so that it is most likely that the formation of dihydroseselin on acid hydrolysis of vellein proceeds through the intermediate osthenol. The sugar present has been shown to be glucose by the partial synthesis of tetra-acetylvellein from osthenol and O-tetra-acetyl α-glucosidyl bromide and hence vellein must be osthenyl β-glucoside.

The absorption spectrum of umbelliferone, observed with a Beckman Model DU spectrophotometer, does not support the observation of Tasaki (1927) that introduction of a hydroxyl group into the benzene nucleus of coumarin reduces the number of absorption maxima in the near ultraviolet region from two to one.

TABLE 1

LIGHT ABSORPTION MAXIMA OF THE COUMARINS FROM V. DISCOPHORA COMPARED WITH THOSE OF UMBELLIFERONE

Compound (in Ethanol)	Maxima					
	mμ	ε	mμ	ε	mμ	ε
Umbelliferone	326	14,950	253	2,425		
Dihydroseselin ..	330	15,190	258	3,143	250	2,484
Osthenol	330	15,490	260	4,881		

Experimental

Melting points are uncorrected unless otherwise indicated.

(i) *Extraction of Vellein.*—Dried and ground stems, leaves, and roots of *V. discophora* (4·6 kg.) were extracted 10 times with ethanol, the extract concentrated to 4·5 l., cooled, and filtered, when a small amount of crystalline material separated. The filtrate was concentrated under reduced pressure to a dark green viscous residue, from which most of the colour was removed by trituration with light petroleum (b.p. 45–90 °C.). The remaining solid and the crystalline material (780 g.) formed long cream coloured prisms of *vellein* (215 g.) from acetone-water (1 : 4 ; 2 l.), m.p. 183–187 °C. From the mother liquor a further quantity of vellein (28 g.), m.p. 187–188·5 °C. was obtained by extraction with ether (total yield : 243 g., 5·3%). On extraction in a soxhlet with acetone, vellein separated from the hot extract, m.p. 187·5–189 °C. (corr.).

Found (on material dried *in vacuo* at 110 °C.) : C, 59·7 ; H, 6·2%.

Calculated for $C_{20}H_{24}O_8.1/2H_2O$: C, 59·8 ; H, 6·3%.

Vellein is optically inactive, soluble in methanol, acetic acid, or hot water, sparingly soluble in cold water, hot ethanol, hot acetone, or hot ethyl acetate, and very slightly soluble in cold ethanol, acetone, or ethyl acetate.

(ii) *Tetra-acetylvellein.*—Vellein (1 g.), acetic anhydride (15 ml.), and pyridine (3 drops) were refluxed for 1 hour, cooled, and poured into water (100 ml.). When the excess acetic anhydride had decomposed, the solid was filtered and washed with sodium carbonate solution, 2N hydrochloric acid, and finally with water. It formed small colourless cubes of *tetra-acetylvellein*, m.p. 193–194 °C., from methanol (120 ml./g.).

Found : C, 59·8 ; H, 5·9 ; CH_3CO, 33·6% ; M, 529.

Calculated for $C_{20}H_{20}O_8.4(CH_3CO)$: C, 60·0 ; H, 5·75 ; CH_3CO, 30·7% ; M, 560·5.

(iii) *Hydrolysis of Tetra-acetylvellein.*—Tetra-acetylvellein (0·4 g.) and alcoholic potassium hydroxide (30 ml. of 2%) were refluxed for 1 hour, cooled, and water (30 ml.) added. After acidifying with hydrochloric acid, neutralizing with potassium bicarbonate, and evaporating to dryness under reduced pressure, the residue was washed with cold water and crystallized from methanol, m.p. 184–187 °C., alone or mixed with vellein.

(iv) *Acid Hydrolysis of Vellein.*—(*a*) Vellein (2·5 g.) and 3·6N sulphuric acid (50 ml.) were refluxed for 1 hour, cooled, and extracted with ether. On evaporation a pale brown oil (1·9 g.) remained. It solidified on cooling and formed stout colourless prisms of dihydroseselin, m.p. 104–105 °C., after several crystallizations from ethanol.

Found : C, 73·0 ; H, 6·6% ; M, 234.

Calculated for $C_{14}H_{14}O_3$: C, 73·0 ; H, 6·1% ; M, 230.

(*b*) Vellein (1 g.) and 0·3N hydrochloric acid (100 ml.) were kept at 80 °C. for 6½ hours, cooled, and filtered, and the solid crystallized from ethanol, m.p. 104–105 °C., alone or mixed with the dihydroseselin obtained above.

Dihydroseselin distilled at 0·05 mm. pressure and a bath temperature of 130 °C., and was recovered unchanged after refluxing for 1½ hours with acetic anhydride and pyridine.

(v) β-*(2,2-Dimethyl-5-hydroxychromanyl)-6-acrylic Acid.*—Dihydroseselin (0·5 g.) was refluxed for 1 hour with potassium hydroxide solution (25 ml. of 10%), cooled, acidified with hydrochloric acid, and the precipitate collected. It formed colourless plates of β-*(2,2-dimethyl-5-hydroxychromanyl)-6-acrylic acid*, m.p. 183 °C. (decomp.), from alcohol.

Found : C, 68·0 ; H, 6·7%.

Calculated for $C_{14}H_{16}O_4$: C, 67·7 ; H, 6·5%.

(vi) *Relactonization of* β-*(2,2-Dimethyl-5-hydroxychromanyl)-6-acrylic Acid.*—β-(2,2-Dimethyl-5-hydroxychromanyl)-6-acrylic acid (0·4 g.) and methanolic sulphuric acid (10 ml. of 5%) were refluxed for 5 hours, cooled, made alkaline with potassium carbonate, and extracted with ether. On evaporation dihydroseselin was obtained as colourless prisms from ethanol, m.p. 104–105 °C. alone or mixed with the material obtained above.

(vii) *Emulsin Hydrolysis of Vellein.*—Vellein (13 g.), water (2 l.), and a solution of emulsin prepared by the method of Mann and Saunders (1940) from 30 almonds were maintained at 45 °C. for 3 hours. The white precipitate (7 g.) was filtered, dissolved in ether, and filtered from a small amount of insoluble material. Removal of the ether and crystallization from benzene yielded colourless prisms of osthenol which appeared to lose solvent of crystallization on standing in air or in a vacuum, leaving a white powder, m.p. 127–128·5 °C.

Found : C, 72·9 ; H, 6·3%.

Calculated for $C_{14}H_{14}O_3$: C, 73·0 ; H, 6·1%.

(viii) *Osthol.*—An ethereal solution of diazomethane (Arndt 1943) was added slowly to osthenol (0·5 g.) in ether (35 ml.) until the evolution of nitrogen had ceased and the solution was a permanent yellow colour. The ether was distilled and the residual oil crystallized. It formed colourless prisms of osthol from aqueous alcohol, m.p. 82–83 °C., undepressed by admixture with an authentic specimen.

Found : C, 73·6 ; H, 6·9%.

Calculated for $C_{15}H_{16}O_3$: C, 73·75 ; H, 6·6%.

(ix) *Dihydro-osthenol.*—Osthenol (0·5 g.) in acetic acid (50 ml.) absorbed 43 ml. of hydrogen (theory requires 48·7 ml. for one double bond) in the presence of a platinum catalyst (Adams, Vorhees, and Shriner 1941) at atmospheric pressure. The product solidified after distillation at 0·03 mm. pressure and formed colourless plates, m.p. 104–105 °C. from aqueous methanol. It depressed the melting point of dihydroseselin to 75–85 °C. Haller and Acree (1934) record 104–106 °C. as the melting point of dihydro-osthenol.

(x) *Conversion of Osthenol to Dihydroseselin.*—Osthenol (0·1 g.) was refluxed for 2 hours with 1·5N hydrochloric acid (10 ml.), cooled, and extracted with ether. The ethereal solution after washing with potassium hydroxide solution, drying, and distillation of the ether left a residue, which formed colourless prisms from ethanol, m.p. 104–105 °C. alone or mixed with dihydroseselin.

(xi) *Partial Synthesis of Tetra-acetylvellein.*—Osthenol (1 g.) and *O*-tetra-acetyl α-glucosidyl bromide (2 g.) (Redemann and Niemann 1942) were dissolved in acetone (25 ml.) and a solution of potassium hydroxide (0·5 g.) in water (10 ml.) added. The solution was allowed to stand at room temperature for 20 hours when the supernatant liquid was decanted, the crystals washed with acetone, filtered (0·25 g.), and recrystallized from methanol forming colourless cubes, m.p. 191–192·5 °C. undepressed by admixture with tetra-acetylvellein.

The authors wish to thank Mr. G. K. Hughes of the Department of Organic Chemistry, University of Sydney, for a sample of osthol, Mr. C. A. Gardner, Western Australian Government Botanist, for collecting and identifying the plant material, the Commonwealth Research Grant to Australian Universities for a grant for equipment, the Drug Panel of the Western Australian Department of Industrial Development for financial support and permission to publish this work, and Dr. Weiler and Dr. Strauss, Oxford, for the microanalyses.

References

ADAMS, R., VOORHEES, V., and SHRINER, R. L. (1941).—"Organic Syntheses." Coll. Vol. 1. 2nd Ed. p. 463. (Wiley & Sons: New York.)

ARNDT, F. (1943).—"Organic Syntheses." Coll. Vol. 2. p. 165. (Wiley & Sons: New York.)

HALLER, H. L., and ACREE, F. (1934).—*J. Amer. Chem. Soc.* **56**: 1389.

MANN, F. G., and SAUNDERS, B. C. (1940).—"Practical Organic Chemistry." p. 365. (Longmans, Green & Co.: London.)

REDEMANN, C. E., and NIEMANN, C. (1942).—*Organic Synth.* **22**: 1.

SPÄTH, E., BOSE, P. K., MATSKE, J., and GUHA, N. C. (1939).—*Ber. dtsch. chem. Ges.* **72**: 821.

SPÄTH, E., and BRUCK, J. (1937).—*Ber. dtsch. chem. Ges.* **70**: 1023.

SPÄTH, E., EITER, K., and MEINHARD, T. (1942).—*Ber. dtsch. chem. Ges.* **75**: 1623.

TASAKI, T. (1927).—*Acta Phytochim. Tokyo* **3**: 21.

WATSON, E. M. (1943-1944).—*J. Roy. Soc. W. Aust.* **30**: 99.

TRANSMISSION-LIKE MULTIPLE BEAM REFLECTION INTERFEROMETRY

By C. F. Bruce*

[*Manuscript received December* 28, 1950]

Summary

The production of transmission-like multiple beam " Fizeau " fringes in reflection, which have high contrast and visibility, is described. These fringes arise, not essentially from the combination of particular reflectivities of the two plates of the interferometer system, but from the elimination of certain rays in the reflected system which differ in phase from other rays by π. The theory of their origin and the experimental conditions required for their production is given. Fringes of equal chromatic order with similar characteristics can also be produced. An application of these fringes to vibration measurement is described where they had important advantages over normal transmission and reflection fringes.

I. Introduction

Tolansky (1948) and co-workers have developed multiple beam interference techniques which have had many important applications in the precise determination of height and depth and in the examinations of surfaces and films.

In these applications " Fizeau " fringes of equal thickness and fringes of equal chromatic order have been used either in transmission or reflection, depending upon the nature of the particular application. Reflection fringes are of particular value in the examination of opaque surfaces, and have other applications, which will be mentioned later, where the use of transmission fringes would be very inconvenient. Reflection fringes are fine dark lines on a broad bright background, and due to this background, monochromatization of the light is very necessary. Absorption in the reflecting films also affects the fringe visibility. Reflection fringes have, therefore, some disadvantages compared with the normal transmission fringes. Though potentially as sharp as the latter, the visibility is dependent very critically on absorption in the reflecting films, and the contrast between the fringes and the background intensity is considerably lower than that for transmission fringes.

Holden (1949) has shown how transmission-like fringes in reflection can arise if particular reflectivities of the two surfaces making up the interferometer system are used.

It is the purpose of this paper to show how transmission-like reflection fringes of high contrast and sharpness can be produced in a different manner, and that their formation is not dependent specifically on combinations of particular reflectivities, but on the elimination of certain beams in the reflected

---

* Division of Metrology, C.S.I.R.O., University Grounds, Sydney.

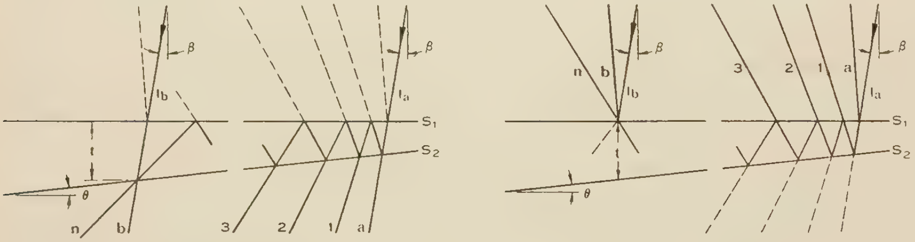
system which differ in phase from other beams by π. Some applications of these transmission-like fringes in which they have advantages over normal transmission and reflection fringes are mentioned.

II. Theory

In order to appreciate the origin of these transmission-like fringes in reflection, brief mention must be made first of conventional multiple beam theory for normal transmission and reflection fringe systems.

(*a*) *Normal Transmission and Reflection*

In Figure 1 let I_a be a beam incident on the two plates S_1, S_2 of an interferometer system giving Fizeau fringes of equal thickness. Portions of this beam undergo multiple internal reflections between the two plates and emerge as a series of beams from plate S_2 forming a transmitted system, and from plate S_1 forming a reflected system. Figure 1 illustrates these two systems and shows full lines for the emergent beams which are used in producing interference and dotted lines for those beams not collected in the particular system used. For clarity, the effects of refraction are not shown.



TRANSMISSION REFLECTION

Fig. 1

In both the transmission and reflection systems it can be seen that there is one emergent beam (labelled *a*) which does not undergo internal reflection. In the transmission system this beam passes through without any reflection. In the reflection system this ray undergoes external reflection at plate S_1. The first emergent beam that has undergone internal reflection is labelled 1, the second is labelled 2, and so on. These emergent beams can combine to give interference and produce localized multiple beam fringes at points whose distance from the plates S_1 and S_2 depend on the angle of incidence β, the wedge angle θ, and the distance between the plates t. The beam I_a is only one of a whole series of incident beams which will contribute to the final interference effects.

For illustrative purposes, interference can be considered for two cases. Interference can occur between the portions of an incident beam I_a which have undergone multiple internal reflections (labelled 1, 2, 3, ... *n*) and the other portion of I_a not undergoing such internal reflections (labelled *a*). Interference can occur between the *n*th emergent ray and the portion of an incident ray I_b not undergoing such internal reflections (labelled *b*).

(i) *Phase Relationships.*—Assuming that the refractive index (μ) of the media between plates S_1 and S_2 is unity, and ignoring terms involving powers of β and θ greater than 2 in the summation, the path differences between ray n and ray a (case (a)), and between ray n and ray b (case (b)) for the two systems are readily shown to be as follows:

(1) *Transmitted system*

Case (a)

$$\delta_{n,a}=2nt-n\beta^2t-2n(n+1)\beta\theta t-\frac{2n(n+1)(2n+1)}{3}\theta^2t.$$

Case (b)

$$\delta_{n,b}=2nt+n\beta^2t-\frac{2n(n+1)(2n+1)}{3}\theta^2t.$$

Expanding these expressions and rearranging terms we get

$$\delta_{n,a}=2nt-n\beta^2t-2n^2\beta\theta t-2n\beta\theta t-\frac{4}{3}n^3\theta^2t-2n^2\theta^2t-\frac{2}{3}n\theta^2t,$$

$$\delta_{n,b}=2nt+n\beta^2t-\frac{4}{3}n^3\theta^2t-2n^2\theta^2t-\frac{2}{3}n\theta^2t.$$

(2) *Reflected system*

Case (a)

$$\delta_{n,a}=2nt-n\beta^2t-2n^2\beta\theta t-\frac{2n(2n^2+1)}{3}\theta^2t.$$

Case (b)

$$\delta_{n,b}=2nt+n\beta^2t-\frac{2n(2n^2+1)}{3}\theta^2t.$$

The expanded forms of these are

$$\delta_{n,a}=2nt-n\beta^2t-2n^2\beta\theta t-\frac{4}{3}n^3\theta^2t-\frac{2}{3}n\theta^2t,$$

$$\delta_{n,b}=2nt+n\beta^2t-\frac{4}{3}n^3\theta^2t-\frac{2}{3}n\theta^2t.$$

It can be seen that if the incident ray is normal to the surface S_1 ($\beta=0$) and if terms involving n^2 and n are considered small compared with terms involving n^3 (n will be of the order of 60), the phase difference in all cases is $2nt-\frac{4}{3}n^3\theta^2t$. Tolansky has used this approximate expression in determining the experimental conditions necessary for high definition multiple beam fringes.

(ii) *Linear Displacement of Beams.*—The linear displacement of the nth beam due to multiple reflections can be determined in a manner similar to that used for the phase relationships.

In the transmitted system the displacement is

$$\begin{aligned}d_n&=2n\beta t+2n(n+1)\theta t,\\&=2n\beta t+2n^2\theta t+2n\theta t.\end{aligned}$$

In the reflected system the displacement is

$$d_n=2n\beta t+2n^2\theta t.$$

Again, if we consider normal incidence and ignore terms in n compared with terms in n^2 we get for both systems

$$d_n=2n^2\theta t,$$

a relation that is also used for investigating the conditions necessary for high definition multiple beam Fizeau fringes.

(iii) *Intensity Relationships.*—This is obtained by the Airy (1831) summation of the amplitudes of all emergent beams contributing to the interference effects. The summation originally carried out by Airy is strictly for Fabry-Perot fringes, but the summation is valid for Fizeau fringes and fringes of equal chromatic order, when certain experimental conditions are fulfilled, as shown by Tolansky.

Let the two surfaces S_1 and S_2 have the same reflection and transmission coefficients. Let t_a and r_a be the amplitude fractions of the incident beam transmitted and reflected respectively at surfaces S_1 and S_2. Let T and R be the corresponding intensity fractions. Let A_T and A_R be the resultant amplitudes transmitted and reflected respectively, and let the resultant phases of these two amplitudes be φ and ψ. Let I_T and I_R be the corresponding intensities. If δ is the phase difference between successive emerging beams

$$\delta=\frac{2\pi}{\lambda}(2\mu t\cos\beta).$$

The resultant amplitudes are given by

$$A_Te^{i\varphi}=t^2+r^2t^2e^{i\delta}+r^4t^2e^{i2\delta}+\ldots.\infty=\frac{t^2}{1-r^2e^{i\delta}},$$

for the transmitted system, and

$$A_Re^{i\psi}=r+(rt^2e^{i\delta}+r^3t^2e^{i2\delta}+\ldots.\infty)e^{i\pi}=r-\frac{rt^2e^{i\delta}}{1-r^2e^{i\delta}}$$

for the reflected system.

The phase factor $e^{i\pi}$ is introduced in the reflected system, since every term in the bracket involves beams undergoing internal reflection, while the term r represents the beam undergoing a single external reflection, and, therefore, differs in phase from the other beams by π.

To derive the intensity relations we have

$$A_Te^{i\varphi}=\frac{t^2}{1-r^2e^{i\delta}}\cdot\frac{1-r^2e^{-i\delta}}{1-r^2e^{-i\delta}},$$

$$=\frac{t^2-t^2r^2e^{-i\delta}}{1-2r^2\cos\delta+r^4}.$$

Equating real and imaginary parts we have

$$A_T\cos\varphi=\frac{t^2-t^2r^2\cos\delta}{1-2r^2\cos\delta+r^4}=\frac{t^2-t^2r^2\cos\delta}{\triangle},$$

$$A_T\sin\varphi=\frac{t^2r^2\sin\delta}{\triangle}.$$

Hence

$$I_T = A_T{}^2 = \frac{t^4(1-2r^2\cos\delta + r^4\cos^2\delta) + t^4r^4\sin^2\delta}{\triangle^2},$$

$$I_T = \frac{t^4}{\triangle}.$$

Similarly for the reflected system we have

$$A_R\cos\psi = r - \frac{rt^2(\cos\delta - r^2)}{\triangle}$$

$$A_R\sin\psi = -\frac{rt^2\sin\delta}{\triangle}.$$

Hence

$$I_R = A_R{}^2 = r^2 - \frac{1}{\triangle}[2r^2t^2(\cos\delta - r^2) - r^2t^4].$$

Substituting intensity terms for amplitude terms we have for the transmitted intensity

$$I_T = \frac{1}{\triangle}T^2. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

The common form of this expression is

$$I_T = \frac{T^2}{(1-R)^2}\;\frac{1}{1+\dfrac{4R}{(1-R)^2}\sin^2\tfrac{1}{2}\delta}, \quad \ldots\ldots\ldots\ldots \quad (2)$$

and for the reflected intensity

$$I_R = R + \frac{1}{\triangle}[RT^2 - 2RT(\cos\delta - R)]. \quad \ldots\ldots\ldots\ldots \quad (3)$$

If we assume no absorption $T+R=1$, the intensity expression reduces to the usual form

$$I_R = \frac{4R\sin^2\frac{1}{2}\delta}{(1-R)^2+4R\sin^2\frac{1}{2}\delta}. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

Holden (1949) considers more fully the intensity relationship by incorporating phase factors in the relations at reflection and transmission, and considering different reflection and transmission coefficients for each of the surfaces S_1 and S_2. For a symmetrical interferometer where both surfaces have the same reflection and transmission coefficients, his full expression reduces to the form

$$I_R = R + \frac{1}{\triangle}[RT^2 + 2TR(\cos\delta + F) - 2R^2T\cos F],$$

where F is a phase factor dependent on the phase changes assumed to take place at transmission and reflection at the surface S_1. This equation is identical with equation (3) if F is given specific values to change the sign of the factor $2TR(\cos\delta + F)$ and $2TR^2\cos F$.

Equations (1-4) show that as R approaches unity (100% reflection) normal transmission fringes will be highly sharpened bright fringes on a dark background, while normal reflection fringes will be highly sharpened dark fringes on a light background.

For no absorption the maximum and minimum values of intensity are

$$I_T^{max}=\frac{T^2}{(1-R)^2} \qquad I_R^{max}=\frac{4R}{(1+R)^2},$$

$$I_T^{min}=\frac{(1-R)^2}{(1+R)^2} \qquad I_R^{min}=0.$$

The fringe shape is defined by the ratio I_{max}/I_{min} and for the transmission fringes this is $(1+R/1-R)^2$. For the reflection fringes the minima become zero irrespective of the value of R. Absorption, however, in practice must be considered. For transmission fringes the ratio I_{max}/I_{min} remains the same, but the effects of absorption for the reflection fringes is important.

Tolansky points out that as a result of absorption the reflection fringe system consists of a classical Airy distribution reduced in intensity by a factor A^2 on which is superimposed a uniform intensity of $R(1-K^2)$, where A is the absorption coefficient of the first surface and $K=1-(A/1-R)$.

Holden shows that minima are raised in intensity by $A^2R/(1-R)^2$ due to absorption, and maxima decreased in intensity by $4AR/(1+R)^2$.

These effects of high background intensity and absorption on the visibility of what are otherwise very sharp reflection fringes are disadvantages, and there is generally lower contrast between fringe and background than for transmission fringes. Monochromatization of the light is, therefore, very necessary when using reflection fringes.

(*b*) *Transmission-like Reflection*

The conversion of normal reflection fringes into transmission-like reflection fringes will occur if all rays reflected externally at surface S_1 are eliminated from the group of internally reflected beams which give rise to interference. These two sets of rays differ in phase by π due to their different conditions of reflection.

In Figure 2 the full line beams indicate those collected to give interference effects, and the dotted line beams indicate those not collected. The important beams eliminated are those labelled a and b, and there will be a whole group of such beams reflected. Interference will then only occur between beams that have undergone multiple internal reflections. All such emergent beams contributing to interference effects have the same phase characteristics, as is so with normal transmission, and consequently reflection fringes with transmission-like characteristics are to be expected under such conditions. In normal reflection fringe systems the interference effects between various multiple reflected beams will give rise to transmission-like effects, but the beams reflected externally are predominant, since the intensity of these beams is very much higher than any of the internally reflected beams. It is therefore to be expected that transmission-like reflection fringes will appear and will show high contrast relative to their background, if the externally reflected beams are sufficiently suppressed. The system is then essentially a transmission one with the first ray to be reflected internally from plate S_2 as the incident beam.

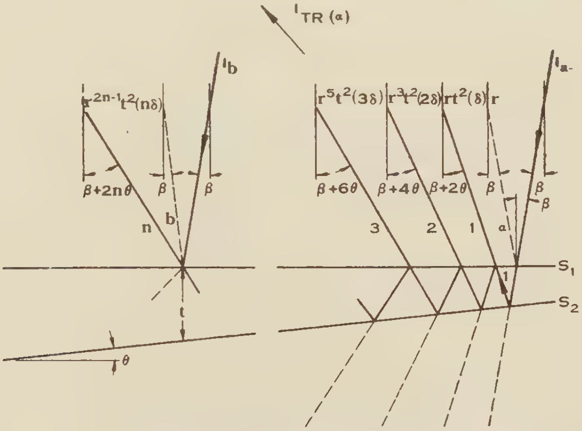
(i) *Intensity Relationships.*—In Figure 2 the elimination of a and b rays means that the intensity distribution is given by the summation

$$
\begin{aligned}
A_{TR}e^{i\alpha} &= rt^2e^{i\delta}+r^3t^2e^{i2\delta}+\ldots\ldots\infty,\\
&= rt^2e^{i\delta}(1+r^2e^{i\delta}+\ldots\ldots\infty),\\
&= \frac{rt^2e^{i\delta}}{1-r^2e^{i\delta}}.
\end{aligned}
$$

By separating real and imaginary parts, squaring, and adding

$$A_{TR}{}^2=I_{TR}=\frac{r^2t^4}{\triangle},$$

where $\triangle=1-2r^2\cos\delta+r^4$.



$I_{TR\,(\alpha)}$
I_b
$r^{2n-1}t^2(n\delta)$
$\beta+2n\theta$
β
β
n
b
t
θ
$r^5t^2(3\delta)$
$r^3t^2(2\delta)$
$rt^2(\delta)$
r
I_a
$\beta+6\theta$
$\beta+4\theta$
$\beta+2\theta$
β
β
β
3
2
1
a
1
S_1
S_2

TRANSMISSION-LIKE REFLECTION

Fig. 2

Using intensity terms this becomes

$$I_{TR}=R\frac{T^2}{\triangle}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

Comparing this relation with equation (1) it is seen that the intensity distribution is identical with that for normal transmission fringes, but reduced in intensity by a factor R, the reflection coefficient of the surfaces S_1 and S_2. In multiple beam conditions this factor approaches unity. Fringes can therefore be expected with transmission-like characteristics without the disadvantages of the effects of background intensity and absorption so prominent in normal reflection fringes. The visibility will approach that for transmission fringes as R approaches unity.

The maximum and minimum intensities for no absorption are

$$I_{TR}{}^{max}=\frac{RT^2}{(1-R)^2},$$

$$I_{TR}{}^{min}=\frac{RT^2}{(1+R)^2}.$$

The ratio
$$\frac{I_{TR}{}^{max}}{I_{TR}{}^{min}}=\left(\frac{1+R}{1-R}\right)^2.$$

Hence the shape of the transmission-like reflection fringes is identical with that for normal transmission fringes.

If there is absorption $T+R+A=1$, instead of $T+R=1$. The intensity relations are

$$I_{TR}{}^{max}=\frac{KRT^2}{(1-R)^2},$$
$$I_{TR}{}^{min}=\frac{KRT^2}{(1+R)^2},$$

where $K=1-\frac{A}{1-R}$.

The fringe shape is therefore not affected by absorption, and so their visibility is not so critically dependent on absorption as with normal reflection fringes.

(ii) *Phase Relationships.*—By simple summation of the path length for the beams between the plates S_1 and S_2, the path difference between the first internally reflected beam emerging and the nth (Fig. 2) is found to be

$$\delta_{1,n}=2(n-1)t-(n-1)\beta^2t-2(n-1)^2\beta\theta t-\frac{2(n-1)(2n^2-4n-3)}{3}\theta^2t.$$

This corresponds to case (a) given for the normal transmission and reflection fringes.

The expanded form of this expression is

$$\delta_{1,n}=2(n-1)t-(n-1)\beta^2t-2(n-1)^2\beta\theta t-\frac{4}{3}n^3\theta^2t+4n^2\theta^2t-\frac{2}{3}n\theta^2t-2\theta^2t.$$

The phase relationships are closely similar to the ones given for normal transmission and reflection, and again the phase lag factor $\frac{4}{3}n^3\theta^2t$ is significant in determining the experimental conditions for the production of the fringes.

(iii) *Linear Displacement of Beams.*—The linear displacement between the 1st and nth emerging beams is

$$d_{1,n}=2(n-1)\beta t+2(n^2-1)\theta t.$$

The factor $2n^2\theta t$ is again of significance. It is clear that if the experimental conditions satisfy the phase conditions for normal transmission and reflection fringes, these conditions will also be satisfactory for transmission-like reflection fringes when the externally reflected beams are eliminated.

III. EXPERIMENTAL

A Fizeau interference system was used as shown in Figure 3. The system was of large aperture and extent for a particular purpose not directly concerned with the present paper. It was particularly valuable in this investigation since the large focal length facilitated not only critical collimation for quite large source apertures, but also the separation of rays at the observing aperture, which differed in inclination to the optic axis by very small amounts. This latter requirement is essential if transmission-like fringes are to be observed. Multiple beam interference studies have generally been done over very small areas of a surface (1 to 2 mm.) which are within the field of a reasonably high-powered microscope, in order to make a precise topographical study of such small areas. The multiple beam effect can, however, be used over extensive areas in many problems giving much greater precision to measurements. For example, the variation in flatness of metal and glass surfaces over a large area can be examined.

The source S was an Osira type mercury discharge lamp, and F is a filter to isolate the mercury green line 5461A. L_1 is a simple condensing lens to concentrate the light on the source aperture in diaphragm D. L_2 is a good achromatic lens of diameter 140 mm. and focal length 1580 mm. N is a large mirror mounted so that it could be turned in a vertical or horizontal plane. The two plates making up the interferometer system were initially two optical flats of high precision and 150 mm. in diameter. These plates could be inclined to one another by suitable amounts to give a convenient number of fringes over the surface. The source aperture and observing aperture in diaphragm D were 7 mm. apart and each was 2 mm. in diameter. A small right-angle prism over the observing aperture enabled the interference pattern to be observed with the naked eye or photographed.

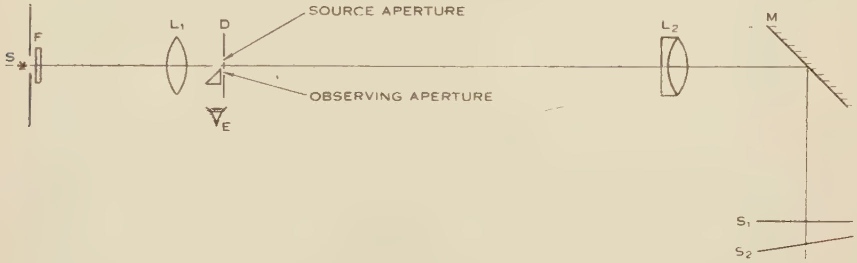


Fig. 3

Plate 1, Figure 1, shows the reflection interference patterns obtained using two 6-in. diameter optical flats for the plates S_1 and S_2, with plate S_2 coated with an opaque layer of silver and plate S_1 coated with silver to give a reflectivity of about 60–65 per cent. The field covered by the photograph is almost the whole surface of the two flats, the extent of the field being limited only by the diameter of 140 mm. of the collimating lens. The surface of plate S_2 was known to be severely distorted, as the flatness of plate S_1 was tested absolutely against a liquid surface as a basic reference of flatness. The same optical system was used, but plate S_2 consisted of a liquid surface of white oil and plate S_1 was not coated with silver in order that the two plates would have comparable reflectivities. Under such conditions two beam Fizeau fringes of good contrast are obtained and are shown in Plate 1, Figure 2. Confirmatory results of the distortion of plate S_2, as shown in Plate 1, Figure 1, were obtained by matching it against a liquid surface. The effectiveness of multiple beam techniques in sharpening the fringes in the optical system described is strikingly shown in Plate 1, Figures 1 and 2. The use of a liquid surface as a standard of flatness in interferometry has been described by Barrell and Marriner (1948).

It is important to see if the conditions for multiple beam interference are reasonably fulfilled in this system. As the diaphragm apertures are slightly off the optical axis the light, after collimation by lens L_1, is not quite normal to the interferometer system. The angle of incidence β will be given by $7/(2\times1580)$ radians, which is approximately 7 minutes. The variation in β due to the size of the aperture is $2\cdot0/(2\times1580)$ radians, or about 2 minutes. The incident angles in the system range from 5 to 9 minutes of angle.

The gap between the two plates could be made any desirable amount. It was always less than 1 mm. For the precise purposes of investigating the collimation, phase, and linear displacement conditions t can always be considered in the range $0\cdot01$ to $1\cdot0$ mm.

The collimation conditions are determined by the relation

$$d\beta=\sqrt{\frac{\lambda dn}{t}},$$

where dn is the change in order produced by a change in incident angle $d\beta$ and t the distance between the interferometer plates S_1 and S_2, and λ the wavelength of the light.

Accepting Tolansky's upper limit for dn of 1/5 the half width W of a fringe we have

$$W=\frac{1-R}{\pi\sqrt{R}}.$$

In this case $R=\sqrt{R_1R_2}$, where R_1 and R_2 are the separate reflectivities of the two plates. The value of R is therefore of the order of 75 to 80 per cent. This gives a fringe width of the order of $\frac{1}{10}$ to $\frac{1}{14}$. Such fringe widths were obtained as shown in Plate 1, Figure 1. For $\lambda=5\cdot5\times10^{-5}$ cm. and $t=1\cdot0$ mm., $d\beta=9$ minutes. For $t=0\cdot01$ mm., $d\beta=1\frac{1}{2}$ degrees. The collimation conditions are therefore satisfactory.

It has been seen that for the transmission-like reflection fringes the important phase lag factor is $\frac{4}{3}n^3\theta^2t$, and this must be small compared with λ. The number of beams (n) combining effectively to give multiple beam interference is probably of the order of 50 for $R=80$ per cent. If we take the usual upper limit for the phase lag factor as $\lambda/2$ we have

$$\frac{4}{3}n^3\theta^2t=\lambda/2.$$

θ is defined by $\theta=\lambda x/2$, where x is the number of fringes per cm. along the wedge of angle θ formed by the two surfaces S_1 and S_2.

Hence $$t=\frac{3}{2n^3\lambda x^2}.$$

Because of the extensive area examined without linear magnification, the values of x are very small, and so the limitations on the value for t are not severe. The values for x ranged from 1·3 to 0·3, corresponding to wedge angles of from about 1 second to 7 seconds. This gives a range for t from 1 to 20 mm. Phase conditions were fully realized therefore in the system.

The linear displacement of the beams is given by the equation

$$d_{1,n}=2(n-1)\beta t+2(n^2-1)\theta t.$$

The only two significant factors in this expression are $2(n-1)\beta t$ and $2n^2\theta t$.

Typical limiting values for the system are given in Table 1.

Table 1

t (mm.)	β (radn.)	θ (radn.)	n	$2(n-1)\beta t$ (mm.)	$2n^2\theta t$ (mm.)	$d_{1,n}$ (mm.)
1·0	$2{\cdot}7\times10^{-3}$	4×10^{-5}	50	0·26	0·20	0·46
0·01	$1{\cdot}5\times10^{-3}$	5×10^{-6}	50	0·15	—	0·15

In the case for $t=0{\cdot}01$ mm., the factor $2n^2\theta t$ is negligible compared with the factor $2(n-1)\beta t$. It is clear, therefore, that linear displacement of beams contributing to interference at any one point shall not visibly distort the interference pattern, and it is therefore an accurate interferogram of the contour of the large area.

When the system is accurately collimated and aligned, a series of images of the source aperture appear on the diaphragm. These arise from reflections from the top and bottom of both plates S_1 and S_2 and from multiple reflections between the plates. All these images can be readily identified, and the distance between particular images on the diaphragm plate can be used to evaluate the wedge angle between the plates S_1 and S_2 and the wedge angle of each of the plates themselves if S_2 is not fully silvered. The particular images of importance are, (i) an image arising from external reflection at the lower surface of plate S_1 (the silvered surface), (ii) an image arising from the first internal reflection from the top surface of plate S_2 and passing through S_1 without any internal reflections at S_1, and (iii) a series of images of gradually decreasing intensity and of uniform spacing, arising from multiple internal reflections between the lower surface of S_1 and the top surface of S_2. From Figure 2 it can be seen that (i) is due to ray a, (ii) is due to ray 1, and (iii) is due to the series of rays 2, 3, ... n. The inclination of any one of the reflected rays differs from its immediate neighbour by 2θ, where θ is the wedge angle of the air film between the plates. As plate S_1 is inclined relative to plate S_2, the images corresponding to rays a and 2, 3, ... n all move much more rapidly than the image corresponding to ray 1.

The motion of the former group of images is due to reflective effects at the lower surface of plate S_1, while the motion of the latter image is due to the small refractive effect only at the lower surface of plate S_1. Of all these images, a is much the brightest and the intensity of the rest decreases from 1, 2, ... n.

Normal reflection fringes are readily obtained by first bringing images due to rays of type 1 into the observing aperture, using mirror M, and then bringing images due to rays of type a, 2, 3, ... n into coincidence and into the observing aperture by tilting plate S_1. The plates S_1 and S_2 should then be closely parallel, and with slight adjustment of plate S_1, normal reflection fringes will appear through the small prism at the observing aperture. The images will then be slightly out of coincidence. The normal reflection fringe can be quickly transformed into the transmission-like reflection fringes by tilting the mirror M very carefully so as to bring the main image slightly away from the observing aperture. Effectively, rays of type a are being eliminated and one can observe the gradual conversion of the fine dark fringes into fine bright fringes.

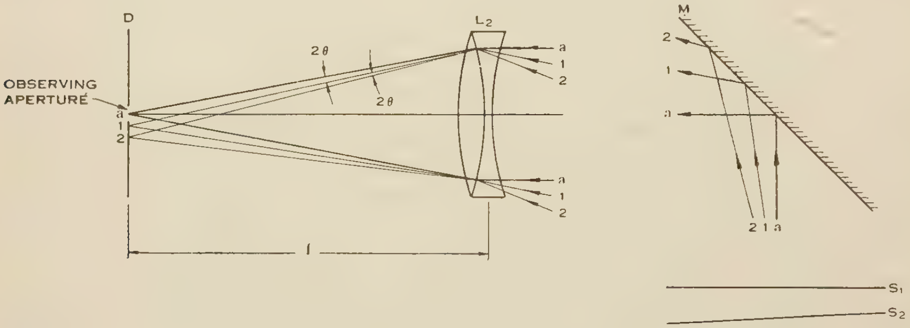


Fig. 4

The transmission-like effect was produced most satisfactorily when the direction of the maximum wedge angle was adjusted so that it was in the same plane as the direction of motion of the mirror. In other words, when the series of multiple images had been arranged in a vertical line on the diaphragm before adjustment to coincidence, a motion of the mirror to give displacement of the image in a vertical direction produced the transmission-like fringes. Further, the direction of motion of the mirror had to be such that the images were displaced in a direction opposite to that in which the multiple images were spread out before adjustment to coincidence. In other words, the direction in which the images have to be displaced depends on the direction of wedge angle. This is clear from Figure 4, where the direction of rays a, 1, and 2 are shown diagrammatically to illustrate the relative positions of the images on the diaphragm D arising from these three rays. Clearly, to eliminate rays of type a, the main image seen through the observing aperture when normal reflection fringes are visible, and which is a partial superposition of all images in the group a, 1, 2, ... n with a predominating, must be moved in the direction opposite to that in which the images due to rays 1, 2, ... n are displaced slightly from the image due to a. These effects are observed experimentally. Further, the displacement of any one image from its neighbour at the diaphragm is $2f\theta$, where f is the focal length of the lens L_2. Consequently, it can be expected that transmission-like fringes will appear with greater definition and contrast for larger wedge angles, since it is then easier to separate a-rays from the rest. This is found to be so experimentally.

Plate 1, Figures 3 and 4, shows the normal reflection fringes and the transmission-like reflection fringes respectively obtained between the two large optical flats, S_1 being coated with silver to a reflectivity of the order of 70 per cent., and S_2 having an opaque layer of silver over the whole surface which already had a film of silver over it which varied in thickness continuously. A strip in the centre free of this first film enabled the film thickness to be evaluated. The sharpness and high contrast of these transmission-like reflection fringes is very evident.

Plate 1, Figures 5 and 6, illustrates the appearance of the transmission-like fringes in unfiltered mercury light, the a-rays being suppressed more in Figure 6 than in Figure 5. Here the wavelengths are resolved as with normal transmission fringes. Such a resolution is not visible with normal reflection fringes due to the high background intensity associated with each wavelength. It becomes possible, therefore, to use light which is not monochromatized in a transmission-like reflection system.

Plate 2, Figures 1 and 2, shows the interferograms obtained for a range of combinations of reflectivities of plates S_1 and S_2. In these, an opaque layer of silver is given a reflectivity of 95 per cent., and a plain glass surface a reflectivity of 4 per cent. It is clear that the production of transmission-like reflection fringes is not at all critically dependent on reflectivities, but that they appear to be best for combinations of reflectivities in the range 60 : 95 to 70 : 95 when the back plate S_2 is fully silvered.

Transmission-like reflection fringes of equal chromatic order were also produced by replacing the mercury source by a high intensity white light source, and inserting a spectroscope and a focusing lens between the eye and the right-angle prism at the observing aperture shown in Figure 3. The focusing lens was used to project the interference fringe system on to the slit of the spectroscope. Since the fringes of equal chromatic order obey the same basic formulae as Fizeau fringes, the characteristics of the transmission-like fringes will be the same as described above.

IV. Applications

Transmission-like fringes have been used in studying the flatness of precision surfaces. In this work these fringes had advantages over the normal reflection fringes because of their higher contrast. A particularly interesting application of these fringes is their use in studying vibrations. Bruce, Macinante, and Kelly (1951) have used multiple beam techniques to determine the amplitudes of vibration of very sensitive vibration detectors. In this work it was necessary to measure amplitudes of the order of $0 \cdot 25\ \mu$. Transmission fringes were first used in these investigations, but reflection fringes were later used, as the reflection system had practical advantages. If, for example, vibrations due to walls are being studied, one plate of the interferometer system will be situated close to the wall, and it would not be possible to mount optical systems on either side of the interferometer plates. The greater intensities in the reflection system also allow shorter exposure times in the photography of the fringes. The amplitude of a sinusoidal vibratory motion with a stationary source of light could be observed as a broadening of the fringe due to one of the interferometer plates moving relatively to the other. Due to the relatively low contrast in reflection fringe systems, the broadened fringes are indistinct. The use of the transmission-like reflection fringes removes this serious difficulty, and Plate 3 shows the appearance of the fringes in mercury green light (5461A) for the stationary and moving state for the transmission, reflection, and transmission-like reflection systems. Obviously, the limit to the amplitude measurable by the broadening of a fringe is one-quarter of a wavelength of light, since broadened fringes would overlap for greater amplitudes. By using an intermittent source of light, however, the amplitude of motion is observed as a slow periodic movement of the sharp fringes, and so greater amplitudes can be measured. In this case also, the transmission-like reflection fringes are more satisfactory than the normal reflection fringes for the measurement of the amplitudes of vibration.

V. Acknowledgments

The author wishes to thank Mr. N. A. Esserman, Chief of the Division of Metrology, for his interest in, and encouragement of this work, and Mr. W. A. F. Cuninghame and Mr. J. C. Kelly, of the Division, for assistance with the experimental work. This assistance included the production photographically of suitable interferograms, and the production of suitable silver coatings on glass by evaporation.

VI. References

Airy, G. B. (1831).—Maths. Tracts. 381.

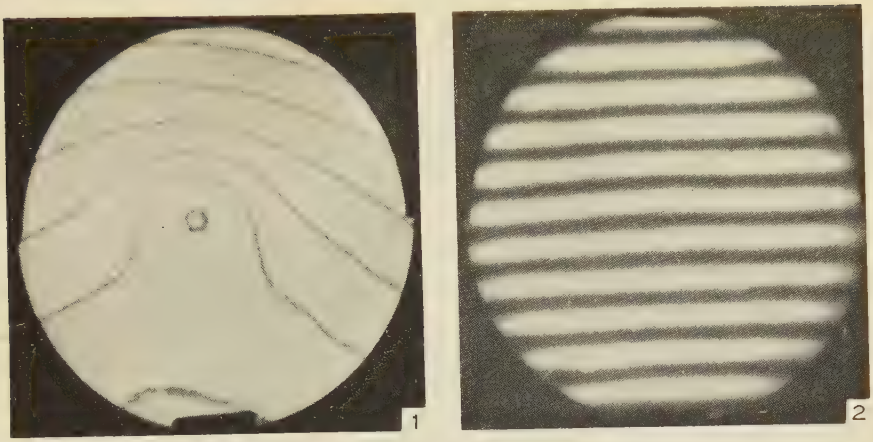
Barrell, H., and Marriner, R. (1948).—*Nature* **162** : 529.

Bruce, C. F., Macinante, J. A., and Kelly, J. C. (1951).—*Nature* **167** : 520.

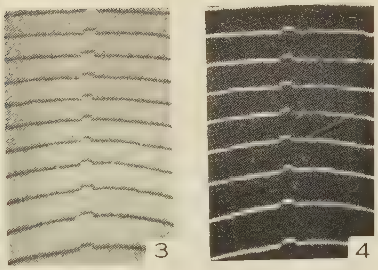
Holden, J. (1949).—*Proc. Phys. Soc.* B **62** : 405-17.

Tolansky, S. (1948).—" Multiple Beam Interferometry of Surfaces and Films." (Clarendon Press : Oxford.)

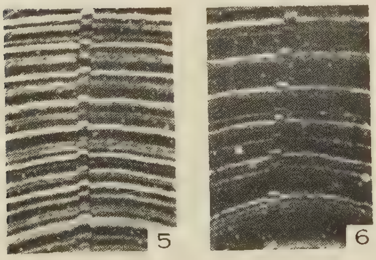
MULTIPLE BEAM REFLECTION INTERFEROMETRY



MONOCHROMATIC LIGHT
REFLECTION

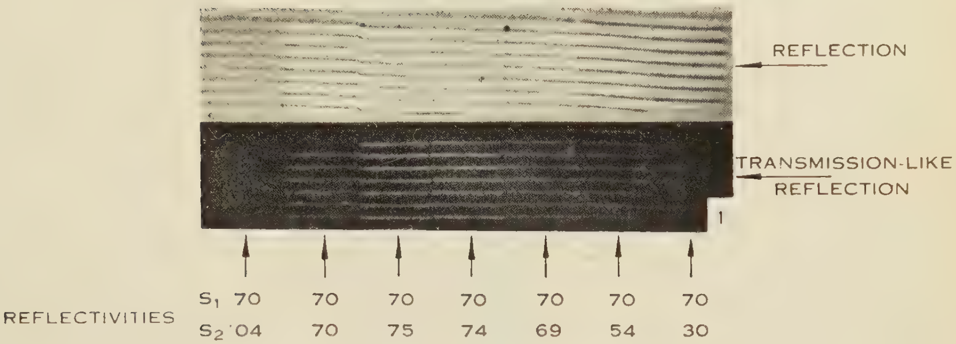

REFLECTION
3
4
TRANSMISSION-LIKE
REFLECTION

MONOCHROMATIC LIGHT

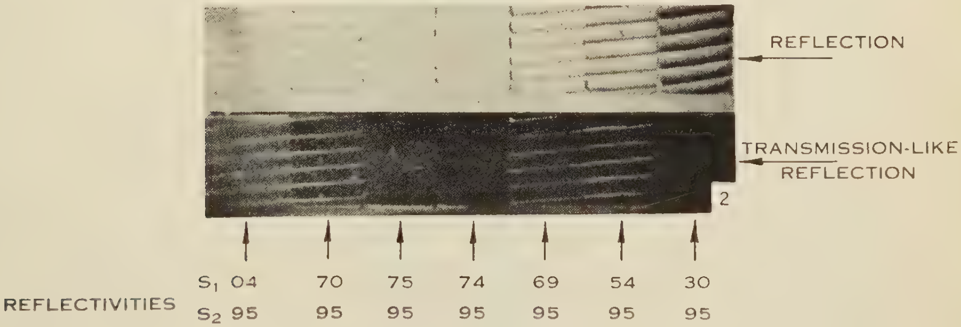


WHITE LIGHT
TRANSMISSION-LIKE REFLECTION

MULTIPLE BEAM REFLECTION INTERFEROMETRY

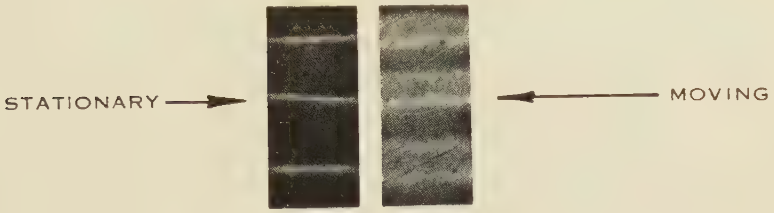

REFLECTION
TRANSMISSION-LIKE REFLECTION
REFLECTIVITIES
S_1 70 70 70 70 70 70 70
S_2 04 70 75 74 69 54 30

1

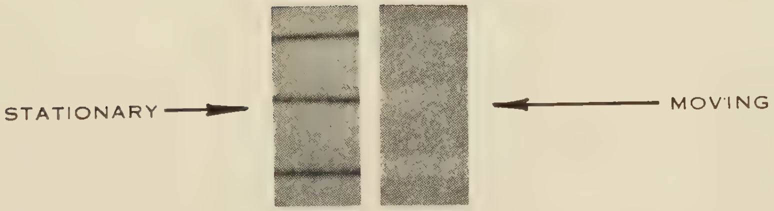

REFLECTION
TRANSMISSION-LIKE REFLECTION
REFLECTIVITIES
S_1 04 70 75 74 69 54 30
S_2 95 95 95 95 95 95 95

2

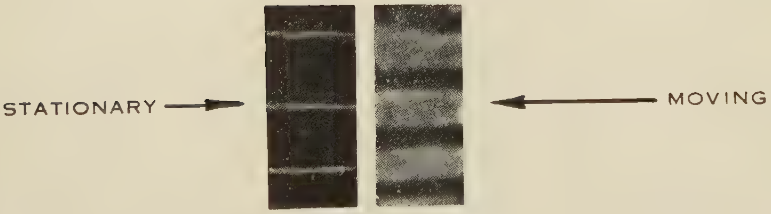
MULTIPLE BEAM REFLECTION INTERFEROMETRY


STATIONARY
MOVING

TRANSMISSION


STATIONARY
MOVING

REFLECTION


STATIONARY
MOVING

TRANSMISSION-LIKE
REFLECTION

SOLAR RADIO-FREQUENCY EMISSION FROM LOCALIZED REGIONS AT VERY HIGH TEMPERATURES

By J. H. Piddington* and H. C. Minnett*

[*Manuscript received November* 22, 1950]

Summary

Solar radio-frequency radiation is analysed into three components ; a basic steady component *B*, a slowly varying component *S*, and various forms of more or less rapid fluctuations called the *X* component. The spectra of all three components are drawn between 600 and 24,000 Mc/s., and suggested extensions to lower frequencies are discussed.

The properties of the *S* component are described in some detail ; these include correlation with sunspot data, polarization, and the location of the sources of origin. Evidence is presented in favour of generation by thermal processes.

It is suggested that the *S* component is due to thermal emission from localized regions at temperatures of about 10^7 °K., often in the vicinity of sunspots. The radiation from a model hot region is examined in detail and the emission spectrum and polarization characteristics are derived. The results are found to compare reasonably with observation.

Thermionic emission of electrons and protons would probably occur from the hot regions. These particles would travel to the Earth with average velocities of a few hundred kilometres per second and may be identical with the slow corpuscular radiation whose presence is deduced from terrestrial magnetic data.

I. Introduction

The observed flux density of solar radio-frequency radiation at frequencies between 600 and 9400 Mc/s. usually varies slowly and shows a marked correlation with sunspot area. In addition to this slow variation, increases of flux density lasting from a few seconds to some hours have been observed (Covington 1948 ; Lehany and Yabsley 1949 ; Minnett and Labrum 1949). It has been shown (Pawsey and Yabsley 1949) that the slowly varying radiation consists of two components ; one is a basic component of thermal origin (Martyn 1948) which originates over the whole of the solar disk and maintains a steady level for periods of at least some years ; the other is a slowly varying component which originates in small " bright " patches observed by Covington (1947) and others (Christiansen, Yabsley, and Mills 1949 ; Piddington and Hindman 1949) and which is roughly proportional to sunspot area (Minnett and Labrum 1949 ; Pawsey and Yabsley 1949). These two components will be referred to as the *B* and *S* components respectively. The more rapid variations of intensity may really consist of several components of different origin and properties, but for the present purposes all such radiation is referred to as the *X* component.

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

The purpose of this paper is to identify the S component and its properties, and to discuss in detail the possibility that it may have a thermal origin in localized regions at very high temperatures.*

II. The Components of Solar Radiation

In order to identify the S component as clearly as possible, the spectra of the three components described above will be computed from experimental data for a wide range of frequencies. The classification of the total radiation into these components is a natural one at frequencies between 600 and 24,000 Mc/s. and has been recognized by Denisse (1949) and Lehany and Yabsley (1949), whose results will be referred to later. At lower frequencies the separation is more difficult and the following discussion applies particularly to frequencies of 600 Mc/s. and higher; however, suggested extensions of the spectra to 85 Mc/s. are also described.

It is found convenient here to express the power flux of solar radiation at a particular frequency in terms of the apparent black-body temperature of the visible solar disk at the frequency concerned. This does not, of course, imply that the origin is necessarily thermal, but merely provides a useful concept and measure of incident flux. The apparent temperature T, corresponding to the total observed flux of solar radiation, consists of three components corresponding to the B, S, and X components listed above. Thus

$$T = T_B + T_S + T_X.$$

(*a*) *The* B *Component*

The origin and properties of the B component have been discussed in some detail by Martyn (1948), Pawsey and Yabsley (1949), and others and need not be reviewed here. The spectrum curve of T_B shown in Figure 1 is based on data from analyses by Pawsey and Yabsley (1949) and Piddington (1950).

(*b*) *The* S *Component*

The slowly varying S component has been recognized (Pawsey and Yabsley 1949) at frequencies of 600 Mc/s. and above. It is in evidence in Figure 2† in which apparent disk temperatures over a period of some months are plotted for frequencies of 200, 600, 1200, and 2800 Mc/s. At the three higher frequencies the day-to-day fluctuations in T correlate strongly with sunspot areas which are also plotted. Excluding small experimental errors the variable part of T is the S component whose further properties will be discussed in Section III. The identification of an S component is more difficult at lower frequencies.

---

* In a paper, received since the present investigation was completed, M. Waldmeier and H. Müller (*Z. Astrophys.* **27**: 58 (1950)) have identified three components of solar radiation. One of these is identical with our S component at a frequency of 10 cm., but at metre wavelengths is identified with "enhanced radiation" (see Section II (*b*) below). Waldmeier and Müller suggest a thermal origin in regions of "coronal condensation" where spectroscopic evidence indicates high electron densities. They do not take account of effects of the spot magnetic field and postulate rather higher electron pressures than those used in the present paper.

† This figure is reproduced from a paper by Lehany and Yabsley (1949) with the addition of 2800 Mc/s. data kindly supplied by A. E. Covington, National Research Council, Canada.

" Enhanced radiation " (Pawsey 1950), which may remain fairly steady for days, sometimes exhibits rather similar characteristics. It may, however, be distinguished from the *S* component in the following ways.

(i) Although high level radiation, at frequencies of 200 Mc/s. and lower, originates near sunspots, large areas of spots may exist without such radiation. This is illustrated in Figure 2 for the period around October 7, when the spot area exceeded 4000-millionths of the disk while the 200 Mc/s. radiation remained at a low level. Another manifestation of the effect is that enhanced radiation usually originates in one, or at most two, regions (Pawsey 1950). The *S* component, on the other hand, correlates much more closely with sunspot area (see Fig. 2) and may have a number of simultaneous sources (Christiansen, Yabsley, and Mills 1949).

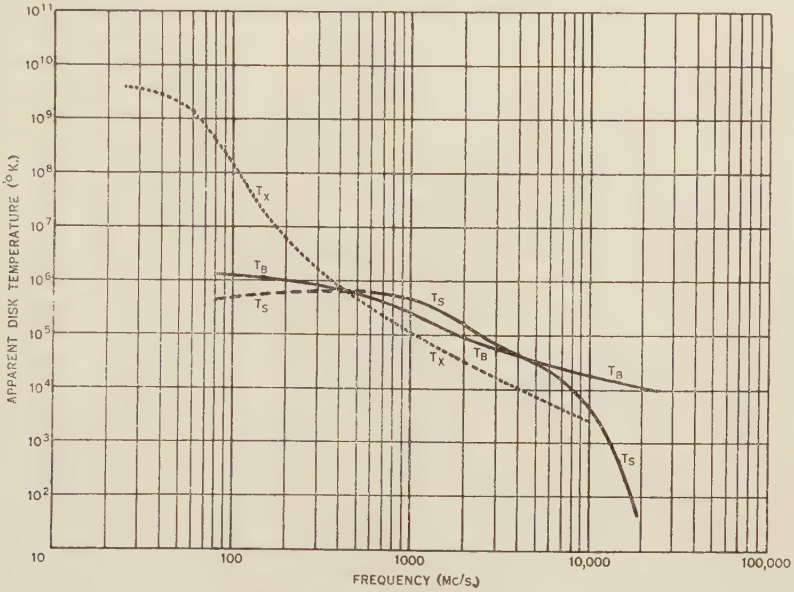


Fig. 1.—The spectra of the three components of solar radio-frequency radiation. That of T_B is constant (at least for several years) in form and level. That of T_S varies in level and the curve shown corresponds to a sunspot area of 5000-millionths. That of T_X is highly variable in both form and level ; the curve shown represents approximate, average, relative values for the various frequencies and for moderately disturbed conditions.

(ii) Enhanced radiation appears to show a correlation with the central meridian passage of large groups of sunspots (Allen 1947 ; Hey, Parsons, and Phillips 1948). T_S, on the other hand, correlates with sunspot area, but otherwise does not show increases of any special significance when the spot group is near the meridian.

(iii) Enhanced radiation is usually strongly circularly polarized (Payne-Scott 1949), while the *S* component is circularly polarized by amounts varying from zero to about 10 per cent.

(iv) T_S never varies by more than a few per cent. in an hour (Section III (*c*)), while enhanced radiation often shows rapid variations (see Fig. 2).

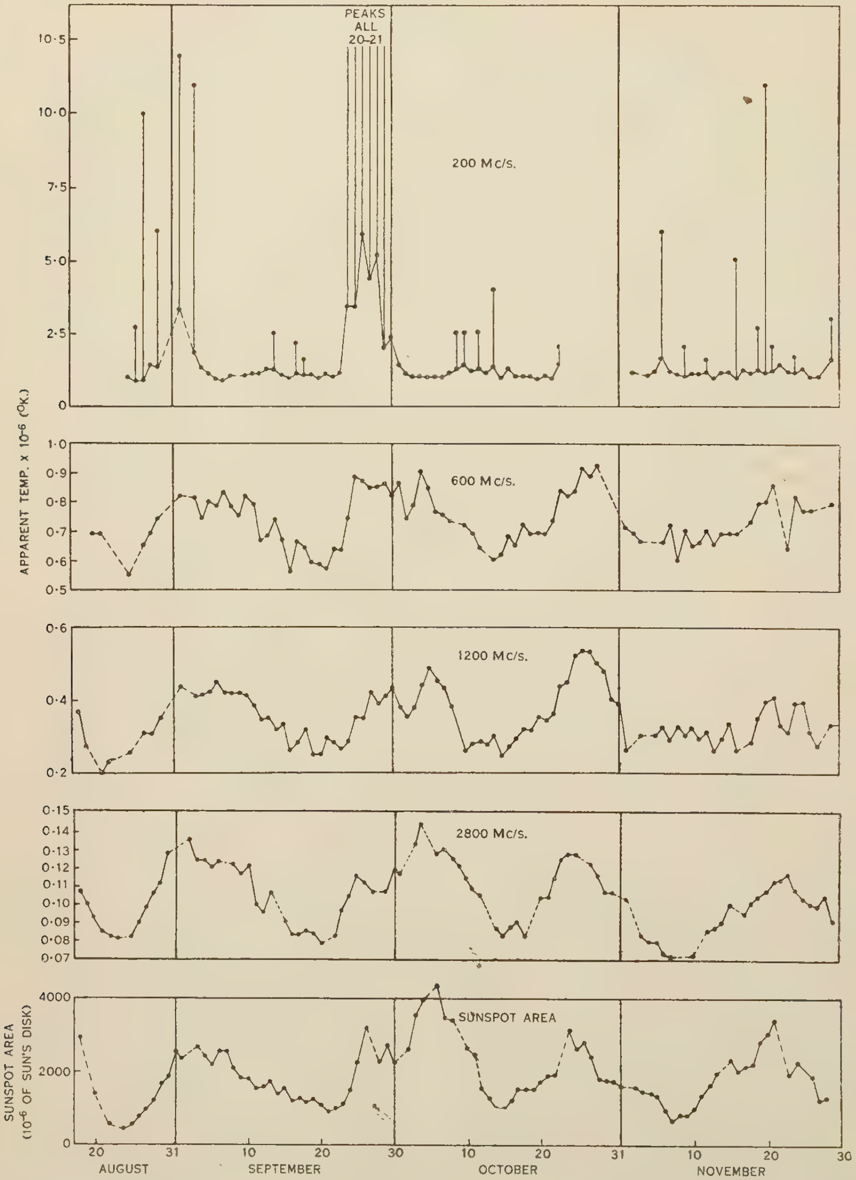


Fig. 2.—The variation with time of the apparent disk temperature at frequencies of 200, 600, 1200, and 2800 Mc/s. together with the projected sunspot area for the period.

(v) The apparent disk temperature of enhanced radiation occasionally rises to very high levels. Values of 10^8 °K. are common, and higher values have been recorded. On the other hand, T_S has not been observed to exceed 10^6 °K. at the higher frequencies where it may be easily identified.

Thus " enhanced radiation " appears to be distinct from the *S* component, but a small but real *S* component may exist at frequencies as low as 80 Mc/s. This and other properties of the *S* component are discussed in Section III.

(*c*) *The* X *Component*

The *X* component includes the phenomena referred to as enhanced radiation, bursts, and outbursts, observed by a number of workers whose results have been summarized by Pawsey (1950). The spectrum of this component is highly variable. The value of T_X at any one frequency, as well as the relative values at different frequencies, varies over a wide range. All that can be indicated are approximate, *average*, relative values at different frequencies. At frequencies between 25 and 200 Mc/s. Appleton and Hey (1946) have given a spectrum curve for a particular occasion of moderate disturbance which agrees in general form with a number of spectra determined by Hey, Parsons, and Phillips (1948). The spectrum from 200 to 24,000 Mc/s. is determined from average relative values of T_X—thus Lehany and Yabsley (1949) found a decrease by a factor of about 140 between 200 and 600 Mc/s. Other observations used were those of Covington (1948), Minnett and Labrum (1949), and Piddington and Minnett (1949). The resultant spectrum is illustrated in Figure 1.

III. OBSERVED PROPERTIES OF THE *S* COMPONENT

The slowly-varying component has been studied by a number of observers for periods ranging from some months to several years in the frequency range from 600 to 24,000 Mc/s. The results are analysed and discussed below and the evidence for existence of the component at lower frequencies is also presented.

(*a*) *Correlation with Sunspot Data*

One of the most outstanding characteristics of the *S* component is its marked correlation with sunspot area. This has been noted by Lehany and Yabsley (1949) at frequencies of 600 and 1200 Mc/s., by Covington (1948) at 2800 Mc/s., and by Minnett and Labrum (1949) at 9400 Mc/s. It is shown very clearly for the three higher frequencies in Figure 2.

Pawsey and Yabsley (1949) have shown that the value of the slowly-varying component is approximately proportional to the total spot area so that

$$T_S = \alpha A,$$

where *A* is the projected area (in units of 10^{-6} of the solar disk) and α is a constant for each frequency. Approximate values of α are given for a number of frequencies in Section III (*b*).

Denisse (1949) determined correlation coefficients between the 2800 Mc/s. intensities observed by Covington (1948) and, successively, (i) relative sunspot numbers, (ii) total sunspot area (not projected area), and (iii) $\Sigma\sqrt{A_1}.H_1$, where H_1 is the maximum magnetic field of a spot whose area is A_1 and the summation was over all visible spots. Correlation coefficients for the 1947 observations were (*a*) 0·53, (*b*) 0·76, and (*c*) 0·87, but the improvement was not so marked for the 1948 results. Denisse considered that three components were present, one corresponding to our *S* component which, he concluded, originated near spots where the electron temperature was $2-8\times10^6$ °K.

(*b*) *The Spectrum of the* S *Component*

We have seen that T_S is related approximately to sunspot area by the equation $T_S=\alpha A$. The values of α over a range of frequencies are used to determine the spectrum of the *S* component. At 600, 1200, and 2800 Mc/s. values of α were found from diagrams published by Pawsey and Yabsley (1949), who plotted the apparent disk temperature against *A* for a large number of observations and determined the straight line best fitting the points.* A value of $\alpha=1\cdot0$ at 9400 Mc/s. was found by Minnett and Labrum (1949), and a value less than $0\cdot06$ may be deduced from the authors' (1949) observations at 24,000 Mc/s.

TABLE 1

INTENSITY OF THE *S* COMPONENT AT FREQUENCIES BETWEEN 80 MC/S. AND 24,000 MC/S.

Frequency (Mc/s.)	Date of Observations	Observers	Value of α (°K. per unit of area)
80	1946–1947	Ryle and Vonberg	110
200	1947	Pawsey and Yabsley	100
600	1947	Lehany and Yabsley	130
1200	1947	Lehany and Yabsley	75
2800	1947	Covington	14
9400	1948–1949	Minnett and Labrum	$0\cdot80$
24000	1948	Piddington and Minnett	$<0\cdot06$

At lower frequencies the *S* component is difficult to identify because of the much stronger *X* component. However, Pawsey and Yabsley (1949) used a statistical treatment in analysing their own 200 Mc/s. data and Ryle and Vonberg's 80 Mc/s. data. They identified two variable components, one of which reached very large values on a small percentage of occasions, while the other was always present, fluctuating by about 50 per cent. of its minimum value. The first of these may be identified with our *X* component. The second "slightly variable" component may be $B+S$ together with experimental errors including errors introduced by the statistical treatment used. Pawsey and Yabsley consider that these errors may account for all the fluctuations, so that values of *S* (or α) derived from their data can only be considered as indicating an upper limit. However, the consistency between values of α found for different periods (when spot areas fluctuated over different ranges of values) suggests that the fluctuations were not wholly experimental. Values of α are derived in Appendix I.

---

* The value for 1200 Mc/s. is doubled, as it is now established (Piddington 1950) that the value of T_B at this frequency is about $2\cdot0\times10^5$ °K. The 3000 Mc/s. value of α is reduced to 14 °K. per unit as this gives a straight line which is a better fit for the larger spot areas.

The values of α over the whole frequency range are collected in Table 1. A curve of T_S for a sunspot area A of 5000 units is plotted in Figure 1 (marked T_S); the uncertain portion at the lower frequencies is dotted. The spectrum of T_S appears to have a maximum at about 1000 Mc/s., falling off to a very low value (not detectable) at 24,000 Mc/s. The spectrum of T_S in units of observed power flux density is shown in Figure 3.

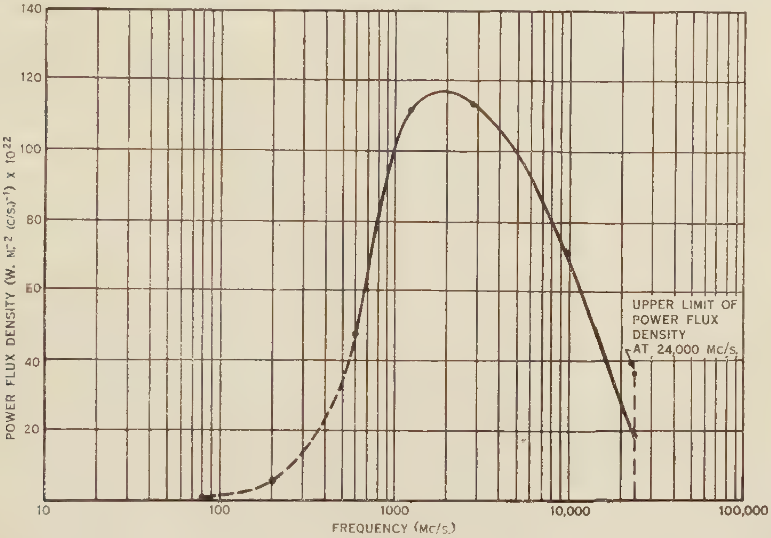


Fig. 3.—The spectrum of the S component expressed in units of power flux density (W.cm.$^{-2}$(c/s.)$^{-1}$) for a sunspot area of 5×10^{-3} of the solar disk.

(c) *Variations with Time*

Figure 2 indicates that T_S varies from day to day by small amounts. Examination of records* of T_S over periods of a number of hours indicates a remarkable steadiness much greater than is necessary to distinguish it from any X components. Piddington and Hindman (1949) have published records of apparent disk temperatures at 3000 Mc/s. for the period October 17, 1948, to March 3, 1949, during which spot areas of more than 6000-millionths were attained on several occasions. Only four times during this period did T_B+T_S vary by as much as 10 to 15 per cent. in periods of a few hours, and even these small variations may have been partly or wholly due to instrumental effects. At no time was any variation discernible in periods of half an hour. The steadiness of T_S is shown in Figure 4, which is a typical plot of T_B+T_S for a 2-hour period when T_S was about 40,000 °K. The zero level (aerial directed away from the Sun) and the level of T_B ($5\cdot4\times10^4$ °K.) are shown as dotted lines. The origin of the small irregularities is at least partly instrumental.

---

* Unpublished observations by the authors at frequencies of 1200, 3000, and 9400 Mc/s.

(*d*) *The Size and Location of Source*

Eclipse observations of Covington (1947), Christiansen, Yabsley, and Mills (1949), and Piddington and Hindman (1949) indicate beyond doubt that the S component originates in small " bright " areas usually, but not always near sunspots.

Covington (1947), while measuring the intensity of radiation at 2800 Mc/s. during the solar eclipse of November 23, 1946, noticed a drop of 25 per cent. of the uneclipsed intensity when a group of spots was covered. The intensity increased by the same amount when they were uncovered. The average apparent disk temperature was $5 \cdot 6 \times 10^4$ °K. (excluding radiation from the vicinity of the spots) and the area emitting the excess radiation was estimated to be $2 \cdot 2$ per cent. of the disk area. These results indicate an equivalent black-body temperature of about $0 \cdot 8 \times 10^6$ °K. for the " bright " area (Covington's own computed value of $1 \cdot 5 \times 10^6$ °K. is apparently in error).

Christiansen, Yabsley, and Mills (1949), making observations during the eclipse of November 1, 1948, at 600 Mc/s., found a number of bright areas contributing a total of about 20 per cent. to the total apparent disk temperature.

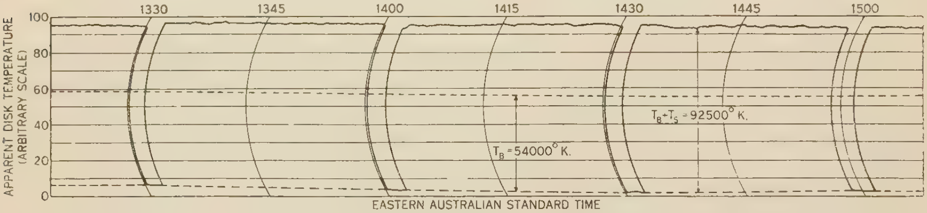


Fig. 4.—The steady and slowly variable components of solar radiation at 3000 Mc/s. over a 2-hour period.

Since the value of T_B at this frequency is $0 \cdot 5 \times 10^6$ °K., the contribution by the bright areas to the apparent disk temperature was about $0 \cdot 1 \times 10^6$ °K. This agrees with the value of T_S deduced from the value of α for 600 Mc/s. (Table 1) and the total sunspot area of 850 units on the day of the eclipse. Three of the eight bright areas located were situated around spot groups, while three other spot groups on the disk were not associated with a bright area. Three other bright areas were close to the positions occupied by spot groups 27 days previously, and one was outside the visible disk and was associated with a prominence. The average equivalent temperature of the bright areas, which averaged 8×10^4 km. in extent, was 5×10^6 °K., the most intense being at least 10^7 °K. Relative changes in the two circularly polarized components of radiation occurred as the areas were covered.

During the eclipse of November 1, 1948, observations were made at 2000 Mc/s. by Piddington and Hindman (1949). They found at least one area with an equivalent temperature of 10^6 °K. Observations by Hagen *et al.* (1948) at 9350 Mc/s. during the total eclipse of May 20, 1947, were not detailed enough to indicate bright areas, but they did show that a larger proportion of radiation originated near the eastern limb of the Sun (last obscured). An examination of sunspot records (1947) for that day showed three large sunspots with fields greater than 3000 gauss, all situated towards the eastern limb.

(*e*) *Polarization of the* S *Component*

Using an aerial which could receive either left- or right-hand polarized radiation, Covington (1949) showed that at 2800 Mc/s. there existed a small circularly polarized component of radiation which varied slowly from day to day. He was able to associate a definite change in the sense of the polarization with the appearance of a large spot group on the solar limb.

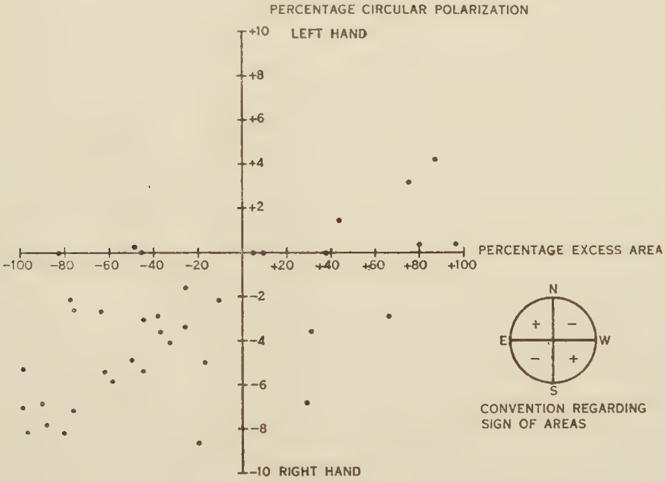


Fig. 5.—Correlation diagram between the excess area in the NE.-SW. quadrants over that in the NW.-SE. quadrants expressed as a percentage of the total area and the percentage circular polarization of the *S* component of radiation.

Similar results were obtained by the present authors during observations made in 1948–49 (unpublished data). It was found that the appearance of a large spot group in the northern hemisphere coincided with an increase in the left-hand* polarized component. The opposite sense was observed for a sunspot in the southern hemisphere. As the spot group approached the central solar meridian, the circularly polarized component reversed sign, and finally vanished when the group reached the limb of the Sun. Since circular polarization was not observed when the slowly-varying component (S) was absent, it may be concluded that this component is responsible for the partial circular polarization of the radiation. When the degree of polarization is expressed as a percentage of the S component, values of up to about 10 per cent. are obtained.

In order to examine the above polarization results more closely, the total sunspot area in the NE.-SW. pair of diagonal quadrants was subtracted from the area in the NW.-SE. pair for each day of Covington's observations (June 1 to August 5, 1948). This excess, expressed as a percentage of the total area, was then plotted against the degree of circular polarization; the latter was expressed as a percentage of the S component by eliminating the B component of aerial temperature which was found from Covington's published results (1948) to be 230 °K. The results of the analysis are shown in Figure 5, from which it will be seen that a sunspot in the NE.-SW. pair of quadrants tends to produce an excess of left-hand circular polarization and the opposite sense for the other pair of quadrants. Only 11 per cent. of the observations disagree with this conclusion, while 14 per cent. are inconclusive; the authors' own measurements, mentioned above, are in general agreement. However, to test conclusively the rule of polarization suggested by the above analysis will require observations extending over a much longer period than those at present available.

---

* In this paper right-hand polarization is defined as clockwise rotation of the electric vector in a fixed plane when observed in the direction of propagation. This is the opposite of that adopted by Covington (personal communication), whose results are accordingly reversed in sense.

It is evident from Figure 5 that over the period concerned there was an average excess of sunspot area in the NW.-SE. quadrants. This appears to correspond with the average excess of right-hand polarization found by Covington.

IV. THE ORIGIN OF THE S COMPONENT

(a) *Theories of Solar Radiation*

Numerous theories have been proposed in explanation of radiation from the disturbed Sun. Kiepenheuer (1946) and Denisse (1947) discuss the emission from electrons spiralling with thermal energies in a magnetic field. They each find an intensity of radio emission in excess of the black-body value corresponding to the electron temperature. This result appears improbable. Other non-thermal or coherent radiation theories (Martyn 1947; Bailey 1949; Haeff 1949) are capable of providing very high emission levels, corresponding to apparent disk temperatures greater than 10^8 °K. However, evidence in favour of a thermal origin for the S component is presented below, and it is shown that these theories are likely to apply to the X component rather than the S.

Theories of the origin of high-level radiation (either S or X component) depending on thermal processes have been proposed by Giovanelli (1948), Ryle (1948), and Lehany and Yabsley (1949). Giovanelli envisages radiation from fast electrons (having velocities equivalent to temperatures up to about 10^{10} °K.) spiralling freely around magnetic lines of force. Discussion of this theory is delayed until Section V (*c*), in which the relevant magneto-ionic theory is introduced.

Ryle (1948) has suggested a thermal origin of the X component in regions where the electron temperature is 10^{10} °K. However, he has over-estimated the electron collision frequency by a factor of about 10^6 and thereby obtained high emission. If suitable values of collision frequency are used (see Section IV (*c*)) the emissivity is found to be negligible. This theory will be referred to in more detail below.

One theory which refers specifically to the S component is that of Lehany and Yabsley (1949), who envisage radiation of the extraordinary magneto-ionic component from the hot corona above sunspots, instead of from the cooler chromosphere. A quantitative investigation of this theory (using their published data) shows that it is not capable of explaining the observational results. For example, consider the flux density of 600 Mc/s. radiation if the extraordinary component originated in the corona for the whole of the disk. Reference to Section V (*b*) and Figure 7 shows that even for spots with strong magnetic fields this would require a spot area of about 25 per cent. of the disk. The ordinary component (of T_B+T_S) would correspond to a temperature of $0{\cdot}5\times10^6$ °K. and the extraordinary component to about 10^6 °K., the mean being $0{\cdot}75\times10^6$ °K. The value of α would therefore be only $\frac{(0{\cdot}75-0{\cdot}50)\times10^6}{0{\cdot}25\times10^6}=1$ compared with the experimental value of 130 (Table 1). A similar argument shows that the theory fails quantitatively at other frequencies.

(*b*) *The Thermal Origin of the* S *Component*

That the S component has a thermal and, incidentally, X a non-thermal origin, is suggested by the following considerations :

(i) It has been shown in Section III (*a*) that T_S varies only very slowly with time, often being quite steady for days and never changing by a detectable amount in an hour. The suggested non-thermal radiating mechanisms (Martyn 1947 ; Bailey 1949 ; Haeff 1949) depend on the rapid movements of streams of gas. Since such movements in the solar atmosphere are very irregular, it is probable that the intensity of radiation would also vary rapidly.

On the other hand, the intensity of thermal radiation would tend to be constant. The heating and cooling of a large volume of gas require time, and the volumes necessary to satisfy the observational data are of the order of 10^{12} cubic kilometres.

(ii) The experimental spectrum curve of the S radiation as shown in Figures 1 and 3 is relatively stable in form, as is shown by the approximate proportionality observed (Pawsey and Yabsley 1949) between the values of T_S on all frequencies and the corresponding sunspot area. On the other hand, the spectrum of T_X, which is represented in Figure 1 as a smooth curve, may vary in form by large amounts, bursts sometimes being observed (Payne-Scott, Yabsley, and Bolton 1947) on one frequency at one time and on another at a different time or not at all.

Any of the mechanisms of non-thermal radiation mentioned above would provide spectra depending critically on a number of factors, including electron densities and density gradients, gas movement, and electric and magnetic field strengths. It is probable that the resultant spectrum would have an irregular and variable form like that of T_X. Thermal radiation, however, even in the presence of a magnetic field, would be expected to exhibit a smooth spectrum curve which would not vary very much for a reasonably wide range of conditions.

(iii) At any given frequency and for a given value of sunspot area the value of T_S is found to lie within a restricted range (about five to one in extreme cases (Pawsey and Yabsley 1949)). However, the value of T_X on a few occasions has reached levels as high as 10^{13} °K. at 60 Mc/s. (Payne-Scott, Yabsley, and Bolton 1947) and $3 \cdot 2 \times 10^6$ °K. at 2800 Mc/s. (Covington 1948). These values of T_X are respectively 10^4 and 10^3 times greater than those given in Figure 1. The highly efficient but complicated mechanism of non-thermal radiation might be expected under suitable conditions to provide almost any level of radiation intensity. Thermal radiation, on the other hand, could not be greater than black-body intensity and, as shown below, is further limited by a decrease in emissivity with increasing temperature.

We may conclude from the above evidence that the S component of solar radiation is probably generated by thermal processes.

(c) *Hot Regions in the Solar Atmosphere*

The eclipse results described above indicate that the S component originates in relatively small " bright " patches, usually associated with sunspots. Additional evidence is provided by the close correlation between sunspot areas and the level of the S component. If the component has a thermal origin we must conclude that the " bright " patches indicate regions where the electron temperature is sometimes 10^7 °K. or more. Spectroscopic evidence of the existence of hot regions in the corona with temperatures up to 6×10^6 °K. has been found by Waldmeier (1947). Electromagnetic processes which would be capable of producing localized volumes of very hot gases have been discussed by Swann (1933), Alfvén (1941), Giovanelli (1947), and others. It is possible that the temperature produced by such heating processes may be limited by the following mechanism to about 10^7 °K. At the temperatures of a few times 10^6 °K., electrons and protons may escape from the gravitational field of the Sun. The presence of a magnetic field would tend to retard this process so that higher temperatures would be necessary. Thus, a process of thermionic emission of the fastest particles may provide an upper limit of temperature (except for short periods) of about 10^7 °K. It is suggestive that the radio data, although scanty, indicate an upper limit of equivalent temperature of the " bright " patches of the same order.

In order to test the theory of thermal origin, the radio-frequency emission spectrum and other properties of hot regions will be calculated, and compared with experimental data. A small and a large model hot region will be considered. Some of the parameters defining the models may be found from experimental data ; others have to be chosen in a rather arbitrary manner but the values selected are considered reasonable.

Suitable lateral dimensions for the model hot regions and their associated sunspots are determined from the eclipse data of Covington (1947) and Christiansen, Yabsley, and Mills (1949). The smaller model region will be considered to have a diameter of 5×10^4 km. and to be associated with a small sunspot having an area about 50-millionths (diameter 10^4 km.) and central field 1500 gauss ; the larger to have a diameter of 2×10^5 km. and to be associated with a spot having an area about 1300-millionths (diameter 5×10^4 km.) and central field 3500 gauss. The vertical dimensions of the hot regions will be taken to be the same as their diameters. This assumption, at least as far as order of magnitude is concerned, appears justified by the very high thermal conductivity along magnetic lines of force and the large scale-height (some 6×10^5 km. at 10^7 °K.). A region of high temperature and pressure having large lateral dimensions would be expected to spread vertically by a large amount. In choosing the level of the base of the hot region above the photosphere the polarization data are most relevant. It is shown in Section V (*f*) that the emission from a hot region will be partially circularly polarized if the region is permeated by a sufficiently strong magnetic field. To explain the observed circular polarization of 3000 Mc/s. radiation, the base of the hot region must extend to within a few thousand kilometres of the photosphere. The exact location of this level, however, does not affect the emission spectrum appreciably (Section V (*e*)).

It remains to determine distributions of electron temperature θ and density N throughout the hot region. In order to simplify subsequent computations and also because our knowledge of the structure of the hot regions is very limited we will consider θ and N as constant throughout the models. In actual fact, N and θ are functions of position and the emission along each ray path is found by integrating emission and absorption for each element of path (Piddington 1950). The total emission from the hot region is then found by a further process of integration over its projected area. The values of the parameters used in the uniform model are, therefore, " average " values for which the computed emission is the same as that from more elaborate models. The assumption of uniformity, although it greatly simplifies calculations, should not appreciably affect the main conclusions of the analysis. A value of θ of at least 10^7 °K. is obviously required to account for the radio results. It is found that an electron density of a few times 10^9 cm.$^{-3}$ is necessary to provide the observed emissivity. The corresponding pressure is rather higher than the average value low in the corona, but perhaps not excessively so, as the heating agency responsible for the unusually high temperatures might be expected to maintain localized regions of high pressure. Values of θ and N are taken then as 10^7 °K. and 2×10^9 cm.$^{-3}$ respectively. The electron collision frequency (which is required later) is given by Piddington (1950) as

$$\nu = 50 N T_e^{-3/2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

It has a value of about 3 sec.$^{-1}$.

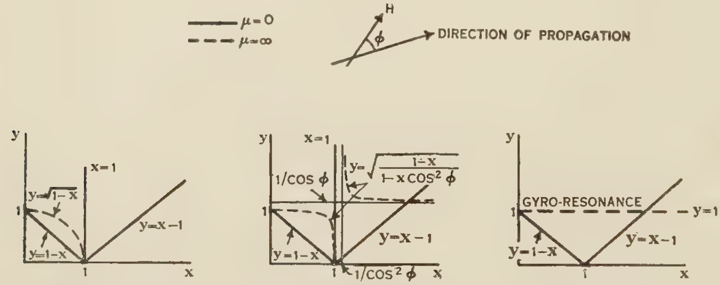
In the following sections the spectrum and polarization of the emission from the two models will be computed.

V. THEORY OF RADIO EMISSION FROM HOT REGIONS

(*a*) *Emission and Absorption by an Ionized Gas*

The thermal emission of radio waves by an ionized medium in a magnetic field may be studied by Appleton's magneto-ionic theory, since the emission is related by Kirchhoff's law to the absorptive power of the medium (Martyn 1948). The theory shows that two modes of propagation (usually known as the ordinary and extraordinary waves) are possible. The propagation characteristics of each are different and in general they are elliptically polarized in opposite senses.

Consider plane waves of angular frequency ω propagating in an ionized gas at an angle φ to the magnetic field H. Under certain conditions the refractive index μ may approach zero or infinity (Ratcliffe 1933). These conditions are illustrated in Figure 6 using the notation $x=4\pi N\varepsilon^2/\omega^2 m$, $y=H\varepsilon/\omega mc$, and $z=\nu/\omega$; ε and m are the electron charge (e.s.u.) and mass respectively, c is the velocity of light, and N and ν have been defined previously. It is assumed that $z<<1$, which is true in the hot region even at very low frequencies ($z=8\times10^{-10}$ at 600 Mc/s. for $\nu=3$ sec.$^{-1}$). Figure 6 shows that there are three conditions for which the refractive index is zero ; these are defined by the relations $x=1$ and $x=1\pm y$. They are independent of the orientation of the magnetic field to the direction of propagation, but in the limiting case of longitudinal propagation ($\varphi=0$) the $x=1$ condition disappears. The refractive index is infinite for the conditions shown by the dotted curves, which are dependent on the orientation of the magnetic field. One condition is associated with the electron gyro-resonance and is defined by $y=1$ if the propagation is longitudinal or if the electron density is low.



(a) TRANSVERSE PROPAGATION $\phi=\pi/2$ (b) OBLIQUE PROPAGATION (c) LONGITUDINAL PROPAGATION

Fig. 6.—The relations between $y=He/\omega mc$ and $x=4\pi N\varepsilon^2/\omega^2 m$ for which the refractive index μ is zero or infinite ; φ is the angle between the field H and the direction of propagation.

The regions where the refractive index for one or other of the modes approaches zero or infinity are deviating regions. The wave is refracted and prevented from propagating past the $\mu=0$ or $\mu=\infty$ boundary except as a highly attenuated evanescent wave. These " reflection " levels prevent the escape of some of the radiation originating in the ionized gas. In the solar atmosphere, y decreases with height above the photosphere and the $x=1-y$ level, which reflects the extraordinary mode, must always lie above the $y=1$ level ; extraordinary wave emission from lower levels cannot pass this barrier. For this reason, the intense $y=1$ gyro-resonance radiation cannot escape (as noted by Ryle (1948) and others). The oppositely polarized mode is reflected, in general, by the $x=1$ level, but for propagation in directions making only a small angle with the magnetic field, reflection occurs at the lower $x=1+y$ level (Ratcliffe 1933). However, in the upper solar atmosphere where ν is very small, the range of angles about the longitudinal direction for which this is true is extremely narrow (Booker 1935). For example, if ν is 3 sec.$^{-1}$, then φ must be much less than 3×10^{3} degrees at a frequency of 600 Mc/s. In general, therefore, appreciable emission in this mode can only escape if it originates in regions above the $x=1$ level, so that regions near the $x=1+y$ level are unlikely to be an important factor in solar radiation.*

The above discussion shows that in estimating the emission from our hypothetical hot region, we need consider only the extraordinary mode radiation from regions above the $x=1-y$ level and the ordinary mode radiation from regions above the $x=1$ level.

(*b*) *Reflection Levels in the Hot Region*

The reflection level for the ordinary mode of propagation occurs when $x=1$. In the model hot region $N=2\times10^9$ cm.$^{-3}$, so that x is only $0 \cdot 44$ at the lowest frequency of 600 Mc/s. (Table 2). The reflection level therefore lies in the denser and cooler gas below the hot region.

In order to estimate the location and shape of the reflection level for the extraordinary mode of propagation, a knowledge of the magnetic field is necessary. It was shown in the previous section that the resonance condition $x=1-y$ is independent of the orientation of the magnetic field. The level is therefore a surface on which the magnitude H of the magnetic field is everywhere constant. The values of H required at the various frequencies are shown in Table 2 for $N=2\times10^9$ cm.$^{-3}$.

TABLE 2

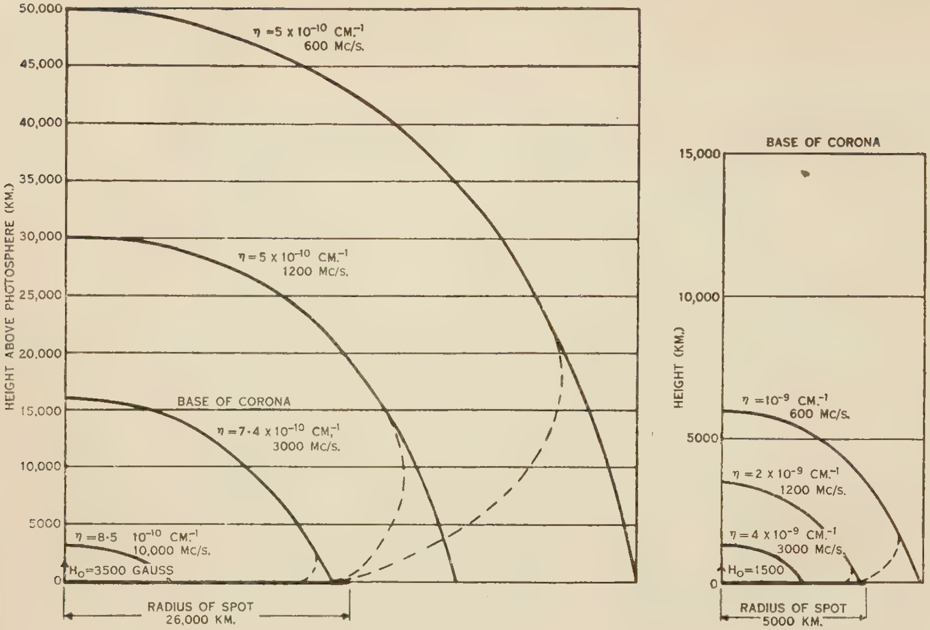
VALUES OF MAGNETIC FIELD SUFFICIENT TO CAUSE REFLECTION

Frequency (Mc/s.)	x	$y=1-x$	Corresponding Value of H (gauss)
600	$0 \cdot 44$	$0 \cdot 6$	120
1200	$1 \cdot 1\times10^{-1}$	$0 \cdot 9$	370
3000	$1 \cdot 8\times10^{-2}$	$1 \cdot 0$	1000
10000	$1 \cdot 6\times10^{-3}$	$1 \cdot 0$	3000
24000	$2 \cdot 8\times10^{-4}$	$1 \cdot 0$	8600

No observations are available on the magnitude and form of the magnetic field in the atmosphere above a sunspot. However, the measurements of sunspot magnetic fields at levels of a few hundred kilometres by Hale and his co-workers have enabled Chapman (1943–44) and Broxon (1949) to construct spot models which should provide reasonable approximations to the fields at higher levels. Chapman's model is a horseshoe magnet beneath the photosphere and Broxon's, a ring of current. Fields at the surface due to each model are in good agreement with observed values except at the edge of the spot. Using Chapman's model,

---

* This appears to invalidate some of the conclusions reached by Saha, Banerjea, and Guha (1947) and by Ryle (1948), who consider that an appreciable quantity of radiation escapes from this level.



(a) 3500 GAUSS SPOT (b) 1500 GAUSS SPOT

Fig. 7.—Sketch of the $y=1-x$ reflection levels for sunspots having a central field H_0 of (*a*) 3500 gauss, (*b*) 1500 gauss. Values of η, representing the approximate gradient of y, are shown for each level.

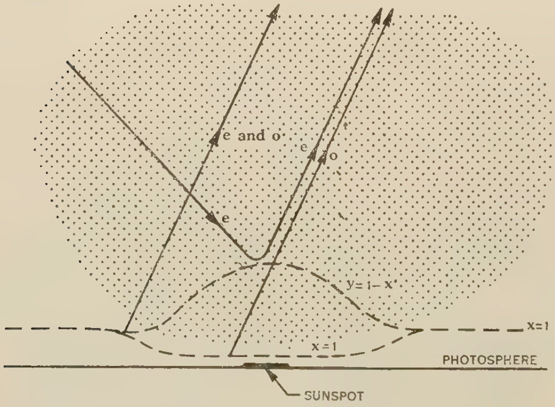


Fig. 8.—Reflection levels and ray paths in a hot region above a unipolar sunspot. *o* and *e* refer to the ordinary and extraordinary waves respectively.

approximate surfaces have been computed over which $x=1-y$ for the frequencies concerned. These have been found for two sizes of sunspot having central fields of 3500 gauss and 1500 gauss respectively, and are sketched in Figure 7. Except for the case of a large spot and a low radio frequency, the radial distance of the shells from the spot centre is less than one spot diameter. Thus although most sunspot groups exhibit bipolar characteristics the field so close to one of the spots will not be appreciably affected by the presence of the second spot.

Figure 8 illustrates roughly the relation of the $x=1$ and $x=1-y$ reflection levels to the region of hot gas above a sunspot. The $x=1$ level is situated in the cooler gas outside the region, while the $x=1-y$ level lies, in general, almost entirely within the region. However, if the sunspot field is weak enough, the $x=1-y$ level for a given frequency may be below the hot region. At 24,000 Mc/s. this is always the case, since the field required for resonance is greater than the maximum values observed in sunspots.

(c) *Absorption Coefficients in the Hot Regions*

The refractive index μ for the ordinary mode and small values of z is equal to $(1-x)^{\frac{1}{2}}$. Since x is small in a hot region (Table 2), the value of μ is approximately unity and the region is non-deviating. The energy absorption coefficient for the ordinary mode under these conditions is given by the Lorentz formula

$$\varkappa=\frac{x\nu}{c}=\frac{4\pi N\varepsilon^2}{\omega^2 m}\cdot\frac{\nu}{c}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

Substituting for ν the value given by equation (1) we have

$$\varkappa=\frac{0\cdot 14N^2}{f^2\theta^{3/2}} \text{ in c.g.s. units,} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

where θ is the electron temperature and f is the wave frequency. It will be observed that $\varkappa$ decreases as θ increases, so that the region becomes more transparent with increase of temperature. Values of $\varkappa$ in the non-deviating region for $\theta=10^7$ °K. and $N=2\times10^9$ cm.$^{-3}$ are shown in Table 3 over the frequency range 600 to 24,000 Mc/s.

The refractive index μ and energy absorption coefficient $\varkappa$ for the extraordinary mode are derived in Appendix II. It is shown that provided the line of propagation lies at an angle not greater than about 45 degrees to the magnetic field :

$$\left.\begin{aligned}\mu&=\left(\frac{1-y-x}{1-y}\right)^{\frac{1}{2}} && \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4)\\ \varkappa&=\frac{\nu x}{c}\cdot\frac{1}{(1-y-x)^{\frac{1}{2}}(1-y)^{3/2}} && \ldots\ldots\ldots\ldots\ldots\ldots \quad (5)\end{aligned}\right\}\text{ for } (1-y-x)(1-y)\gg xz.$$

The variations of μ and $\varkappa$ with y for a small constant value of x are shown in Figure 9. Over a large range the refractive index is almost unity and $\varkappa$ is approximately equal to $\nu x/c$. In this non-deviating region, therefore, $\varkappa$ is the same for both modes and has the values given in Table 3.

The absorption coefficient becomes relatively large near the reflection level $y=1-x$ and near $y=1$ (gyro-magnetic resonance). In the neighbourhood of these points the expressions given above are invalid, the true values of $\varkappa$ being shown in Figure 9. However, the extraordinary mode cannot propagate beyond the reflection level $y=1-x$, so that only absorption up to this point is of interest. Values of $\varkappa$ for $y=1-x$ are given in Table 3.

TABLE 3

ABSORPTION COEFFICIENTS

Frequency (Mc/s.)	$\varkappa$ For Both Modes in Non-deviating Region (cm.$^{-1}$)	$\varkappa$ For Extraordinary Mode at Reflection Level $y=1-x$
600	$4\cdot5\times10^{-11}$	$2\cdot1\times10^{-5}$
1200	$1\cdot1\times10^{-11}$	$4\cdot2\times10^{-5}$
3000	$1\cdot8\times10^{-12}$	$1\cdot1\times10^{-4}$
10000	$1\cdot6\times10^{-13}$	$3\cdot5\times10^{-4}$
24000	$2\cdot8\times10^{-14}$	$8\cdot5\times10^{-3}$

(*d*) *Optical Depth*

It is convenient to express the radiation from the hot region in a given direction in terms of the optical depth of the region. This is equal to the value of $\int \varkappa dl$ determined by retracing the path of rays emerging in the given direction, where $\varkappa$ is the energy absorption coefficient of the mode considered and dl is an element of path.

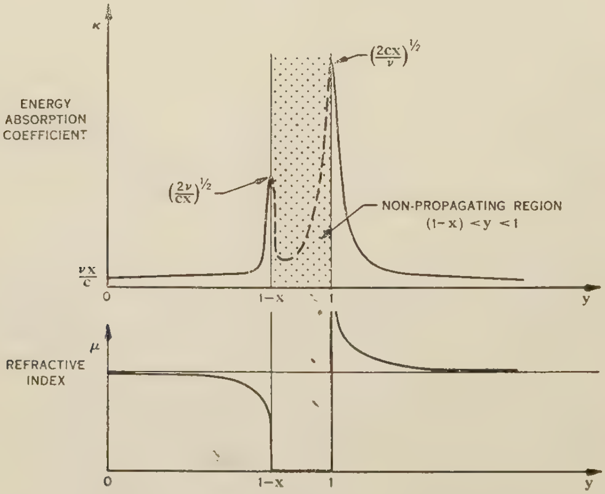


Fig. 9.—The variation with y of the absorption coefficient $\varkappa$ and refractive index μ for a small constant value of x.

For the ordinary mode, the region is everywhere non-deviating and the absorption coefficient is constant and equal to $\nu x/c$ (Table 3). Thus the optical depth for the ordinary mode is equal to $(\nu x/c)l$, where l is the length of the ray path through the region (Fig. 8). Many of the ray paths of the extraordinary mode also traverse the hot region without deviation at the reflection level $y=1-x$. For such rays the optical depth is equal to that of the ordinary mode.

The optical depth of the hot region, however, is much greater over an area approximately equal to the projected area of the $y=1-x$ reflection level. This is because the extraordinary-mode rays emerging over this area have been reflected in the highly absorbing layers adjacent to the reflection level, and the total length of ray path is also greater as shown in Figure 8. For the relatively limited region of high absorption near the level $y=1-x$, it is reasonable to assume that y varies linearly with the distance s from the reflection level. Thus $y=y_0-\eta s$, where η is a constant and y_0 is the value of y at $s=0$; the value of x is constant throughout the region. Substituting in equation (5), and bearing in mind that $y_0=1-x$, we have

$$\varkappa=\frac{\nu}{c}\,\frac{x}{\eta^{\frac{1}{2}}s^{\frac{1}{2}}(x+\eta s)^{3/2}}$$

for $\eta s(x+\eta s)>>xz$.

For small values of s the equation is valid provided $\eta s>>z$, and for large values of s it remains valid until y begins to depart from the linearity assumed. The lower and upper limits of s will be called s_1 and s_2 respectively.*

The optical depth of a layer lying between these two limits (remembering that a ray traverses the layer twice) is given by

$$\begin{aligned}2\int_{s_1}^{s_2}\varkappa ds&=\frac{2\nu x}{c\eta^{\frac{1}{2}}}\int_{s_1}^{s_2}\frac{ds}{s^{\frac{1}{2}}(x+\eta s)^{3/2}},\\&=\frac{4\nu}{c\eta}\left[\frac{1}{\left(1+\dfrac{x}{\eta s_2}\right)^{\frac{1}{2}}}-\frac{1}{\left(1+\dfrac{x}{\eta s_1}\right)^{\frac{1}{2}}}\right].\end{aligned}$$

Provided $\eta s_1<<x$, the second term in this expression may be neglected. As explained above it is necessary that $\eta s_1>>z$, but since $z<<x$ in the hot region (z/x has values between 10^{-7} and 2×10^{-9}), it is possible to choose s_1 so that both inequalities are satisfied. The value of s_2 is restricted by the condition that η has been taken as a constant in order to simplify the integration process. An examination of Figure 7 suggests that if $\eta s_2 \not> 0\cdot2$ this condition is approximately satisfied. The optical depth of the deviating layer is then given by $\frac{4\nu}{c\eta}(1+5x)^{-\frac{1}{2}}$ which, although slightly lower than the true value because of the limitation of the integration range, is sufficient for the present purposes. Values of η are shown in Figure 7, falling within the range 5×10^{-10} cm.$^{-1}$ to 5×10^{-9} cm.$^{-1}$. Values of optical depth range from a maximum of $0\cdot64$ for 1200 Mc/s. radiation from the larger sunspot to zero for 24000 Mc/s. radiation from either spot.

In the above calculations, Doppler and relativistic effects resulting from the high thermal velocities of the particles have not been considered. The authors have made preliminary investigation of these effects, and, contrary to Ryle (1948), believe that the former may play a part in determining emission. However, even if this is shown to be the case it will not greatly affect the quantitative results given in Tables 4 and 5.

---

* This treatment is similar to that of Ryle (1948), who further approximates the expression for $\varkappa$ and restricts his limits of integration; he finds an optical depth smaller by a factor of 10.

It is convenient here to mention briefly Giovanelli's (1948) theory of emission by fast electrons spiralling about magnetic lines of force. The emission level referred to is apparently the $y=1$ level from which radiation is emitted efficiently even in the absence of collisions. However, as has been shown above, this radiation is reflected, and partly absorbed, by the $y=1-x$ level which lies above the $y=1$ level. This process is effective even for very small values of x. Radiation from the $y=1-x$ level is only available if the electrons suffer collisions. It may be possible, as mentioned above, for radiation from the $y=1$ level to escape because of the Doppler frequency shift, but such a mechanism requires further investigation.

(*e*) *Emission Spectra*

Let the optical depth of a hot region be τ_0 for the ordinary mode and τ_E for the extraordinary mode. The equivalent black-body temperature for the region, for an aerial which receives both modes with equal efficiency is given by

$$T_e=\frac{\theta}{2}(1-e^{-\tau_o})+\frac{\theta}{2}(1-e^{-\tau_E}),$$

where $\tau_0=(\nu x/c)l$, l being the length of the ray path in the hot region. The value of τ_E is the same as that for τ_0 for all rays which do not undergo reflection at the $y=1-x$ level. Such rays (illustrated in Fig. 8) originate in what we will call the "transparent zone" of the hot region. The optical depth for rays propagated in the extraordinary mode which do encounter the $y=1-x$ reflecting shell (see Fig. 8) is given by

TABLE 4

EXCESS EQUIVALENT TEMPERATURES (°K.)

Frequency (Mc/s.)	"Transparent" Zone		"Bright" Zone	
	Large Model	Small Model	Large Model	Small Model
600	$5{\cdot}9\times10^6$	$2{\cdot}0\times10^6$	$7{\cdot}5\times10^6$	$3{\cdot}4\times10^6$
1200	$2{\cdot}0\times10^6$	$5{\cdot}6\times10^5$	$4{\cdot}3\times10^6$	$1{\cdot}5\times10^6$
3000	$3{\cdot}6\times10^5$	$9{\cdot}0\times10^4$	$2{\cdot}4\times10^6$	$6{\cdot}1\times10^5$
10000	$3{\cdot}2\times10^4$	$8{\cdot}0\times10^3$	$1{\cdot}9\times10^6$	—
24000	$5{\cdot}6\times10^3$	$1{\cdot}4\times10^3$	—	—

$$\tau_E=\frac{2\nu x}{c}\cdot l+\frac{4\nu}{c\eta}(1+5x)^{-\frac{1}{2}}.$$

Such rays are said to emerge from the "bright zone" of the hot region. The values of l for the two model regions have been given as 5×10^4 and

2×10^5 km. Corresponding values of equivalent temperature for the transparent and bright zones are given in Table 4.

In the above treatment no account has been taken of any change in the B component of solar radiation due to the presence of the hot regions. At low frequencies the equivalent temperature of the hot region is much larger than T_B so that changes in the latter may be neglected. At high frequencies the hot regions are practically transparent so that radiation from below them is not appreciably absorbed. We will assume, for lack of more detailed knowledge, that the equivalent temperature of the lower region is equal to T_B. Thus, for the present purposes, the figures in Table 4 will be considered to apply to the S component only.

In order to compare these results with those of observations of the S component, the constant α (see Section III (b)) is calculated for each model and for all frequencies. We have

$$\alpha = \frac{T_e}{10^6} \cdot \frac{\text{projected area of hot region}}{\text{projected sunspot areas}}.$$

The model diameters are 2×10^5 and 5×10^4 km. and their spot diameters 5×10^4 and 1×10^4 km. respectively. Diameters of the bright zones may be determined from the sketches of the $y = 1 - x$ reflection levels given in Figure 7. Table 5 gives the values of α which were obtained, together with the observed values reproduced from Table 1.

TABLE 5

VALUES OF α (°K. PER 10^{-6} OF SOLAR DISK)

Frequency (Mc/s.)	Observed Value	Large Model			Small Model		
		"Transparent" Zone	"Bright" Zone	Total	"Transparent" Zone	"Bright" Zone	Total
600	130	77	23	100	46	5·3	51
1200	75	29	6·5	36	13	1·3	14
3000	14	5·4	2·2	7·6	2·2	0·19*	2·4
10000	0·80	0·51	0·28	0·79	0·20	Nil	0·20
24000	<0·06	0·09	Nil	0·09	0·04	Nil	0·04

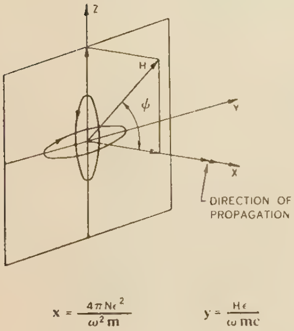
* Provided the hot region extends down to the low level of this reflecting shell.

(f) *Polarization*

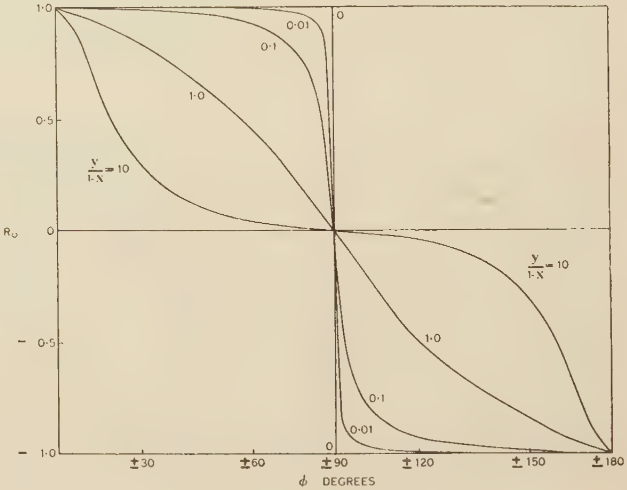
The magnetic vectors of the two component waves which can travel in a magneto-ionic medium are transverse to the wave normal and describe ellipses with opposite senses of rotation. For the low collision frequencies considered here, the two polarization ellipses are located as shown in Figure 10 (a) with respect to the wave normal and the direction of the static magnetic field H.

Following Ratcliffe (1933), the polarization of a component wave will be denoted by R, the ratio of the ellipse axis lying in the plane ZOX to the axis at right angles to this plane. The sense of the rotation, as defined in Section III (*e*), is left- or right-handed according as R is positive or negative. The wave is linearly polarized if R is zero and circularly polarized if $R=\pm 1$.

The polarization of each wave changes progressively as the characteristics of the medium vary along the wave path, and at each point depends only on the quantities x, y, and φ defined in Section V (*a*). If the polarizations of the ordinary and extraordinary waves are R_0 and R_e respectively, then $R_0 R_e=-1$; the variation of R_e with φ is shown in Figure 10 (*b*) for various values of the parameter $y/(1-x)$. Only positive values of this parameter are shown since $x<1$ in all cases of interest. It will be observed that the sign of R_0 (and hence R_e) reverses, if the angle φ between the wave normal and the magnetic field passes through 90 degrees; at this value of φ both waves are linearly polarized. In general, the polarization of each wave is elliptical but becomes circular for low values of $y/(1-x)$.



(a)



(b)

Fig. 10 (*a*).—The relation between the polarization ellipses, the static magnetic field, and the direction of propagation.

Fig. 10 (*b*).—The polarizations of the ordinary and extraordinary waves, R_0 and R_e respectively, as a function of the angle φ between the static magnetic field and the direction of propagation for various values of $y/(1-x)$.

If either x or y vanish, the polarizations of the waves remain unchanged thereafter at the values appropriate to the values of $y/(1-x)$ and φ at the "boundary" of the medium. In the cases of interest here both x and y decrease gradually towards the outer parts of the solar atmosphere and $y/(1-x)$ tends towards zero. The limiting polarization of the emergent wave components is therefore circular and the sense depends on whether φ is acute or obtuse in the boundary regions.

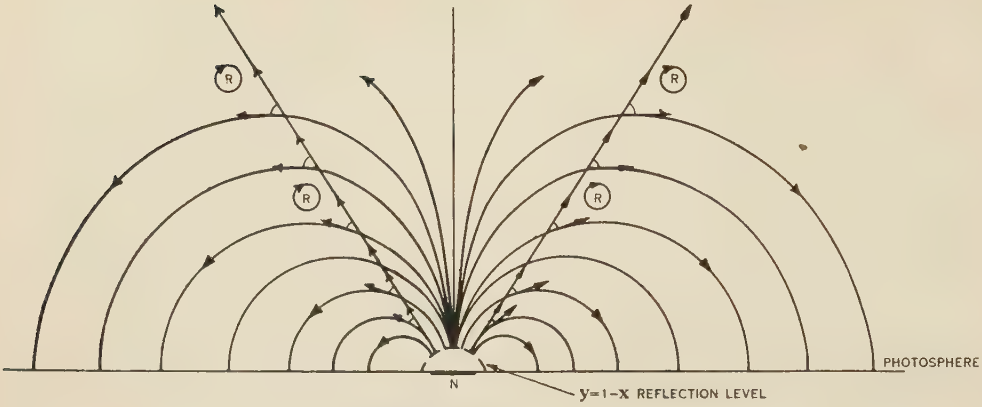
If the two components of the radiation are equally intense, the resultant limiting polarization is random ; if not, a proportion of the radiation is circularly polarized. Radiation from the non-deviating parts of a hot region in the solar atmosphere is equally intense for both modes of propagation and is therefore randomly polarized. However, as shown in Section V (*e*), the " bright " zone associated with the $y=1-x$ shell is an intense source of the extraordinary wave component and there will therefore be an excess of circularly polarized radiation.

Figure 11 shows the probable distribution of the magnetic field of (*a*) a unipolar sunspot, (*b*) a bipolar group. It is assumed that neighbouring sunspots are remote enough not to affect the field to an appreciable extent and distortions due to a general solar magnetic field are also assumed to be negligible. Consider the extraordinary wave emitted by the " bright " shell of the unipolar spot and travelling along the ray path shown. The angle φ changes progressively along the path and approaches a limiting value at great distances from the sunspot. If the lines of magnetic force re-enter the photosphere perpendicularly as shown, the limiting angle cannot exceed 90 degrees for any ray direction. The sense of the limiting circular polarization is therefore independent of the direction from which the hot region is observed and is right-handed if the spot has northern polarity.

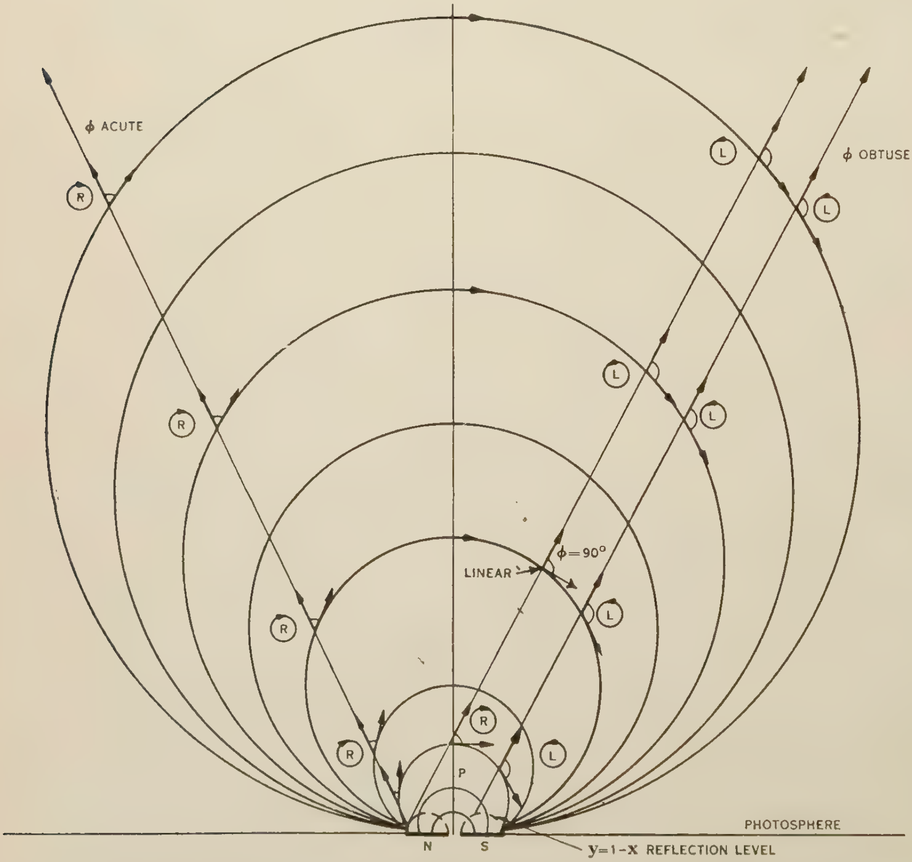
At short distances from a bipolar spot, the extraordinary wave emission from one " bright " shell is elliptically polarized in the opposite sense to that from the other. This is shown in Figure 11 (*b*) by the fact that φ is acute in one case and obtuse in the other. As the wave travels away from the spot, the lesser value of φ increases through 90 degrees and both values converge towards the same limit. The limiting polarization of the emission from both shells is therefore circular and in the same sense. For ray directions on the other side of the plane of symmetry *PP*, the sense of the limiting polarization is reversed since the limiting value of φ passes through 90 degrees for directions parallel to this plane. The sense is right-handed for directions on the northern side of the plane and the opposite on the other (Figure 11 (*b*)).

Bipolar groups are the most common type of sunspot. Although the form of the group is often complex, the line joining the two poles is usually inclined at only a small angle to the solar equator. The plane of symmetry of the group therefore lies approximately in a solar meridian plane. During the present sunspot cycle, the leading spot of a bipolar group in the northern hemisphere is a south pole and the opposite in the southern hemisphere. Thus Figure 11 (*b*) shows that if the group is in the NE. or SE. quadrants, the excess circular polarization should be left-handed. The sense should be right-handed for groups in the other pair of diagonal quadrants. This result is in qualitative agreement with the results described in Section III (*e*).

The percentage circular polarization of the *S* component may be estimated roughly from Table 5. At 3000 Mc/s. the extraordinary wave emission from the " bright " shell contributes from 8 to 29 per cent. of the total radiation, depending on the size of the sunspot. Thus the percentage circular polarization should be from 8 to 29 per cent. of the difference between the percentage sunspot areas in


R
R
R
R
PHOTOSPHERE
N
y=1-x REFLECTION LEVEL

(a) UNIPOLAR SUNSPOT - NORTHERN POLARITY


φ ACUTE
R
L
φ OBTUSE
L
L
R
L
φ=90°
LINEAR
R
L
R
R
L
P
PHOTOSPHERE
N
S
y=1-x REFLECTION LEVEL

(b) BIPOLAR SUNSPOT

Fig. 11.—The probable distribution of the magnetic field above (*a*) a unipolar sunspot, (*b*) a bipolar group, showing the variation of the angle φ along various ray paths of the extraordinary wave.

the two pairs of diagonal quadrants. The results shown in Figure 5, while not inconsistent with this estimate, are too scattered to allow definite conclusions to be drawn.

VI. Discussion

The investigation described in the previous section shows that it is reasonable that the slowly variable component should have a thermal origin in regions at temperatures of about 10^7 °K. The emission from such regions would correlate with sunspot area, and would vary slowly because of the large volume of gas to be heated and cooled.

An examination of Table 5 shows that the theoretical and experimental values of α are generally similar, the latter being somewhat larger. It is clear that slight adjustments to the various parameters of the models, notably to the value of electron density, would permit a much closer fit to be obtained. Such a procedure is not warranted, as it is only intended to show that general quantitative agreement can be obtained. At lower frequencies than those considered, the models become largely opaque, so that α does not increase much further. Furthermore, at a frequency of about 400 Mc/s. (for the models used here) the $x=1$ level rises to the top of the hot region and emission falls to a very low value. In actual practice the electron density is not uniform and the decrease in emission with frequency will be more gradual. This is in qualitative agreement with the experimental curve of T_S in Figure 1.

The polarization theory is in qualitative agreement with the analysis of observations given in Section III (*e*). The rule relating the sense of the excess circular polarization to the net sunspot area in diagonal quadrants of the solar disk should reverse after the next minimum of the sunspot cycle.

Hot regions may exist independently of sunspots, which would account for the bright areas remote from sunspots observed during the eclipse of November 1, 1948 (Christiansen, Yabsley, and Mills 1949). Table 5 shows that the 600 Mc/s. emission from hot regions is still substantial in the absence of a magnetic field. The radiation from such a region would be randomly polarized, but insufficient data from eclipse observations are available to test this conclusion.

It is possible, as mentioned in Section IV (*e*), for electrons and protons in equal numbers to escape from the Sun's gravitational field at temperatures of a few times 10^6 °K. Such corpuscular radiation from hot regions at temperatures of 10^7 °K. would travel to the Earth with a velocity of a few hundred kilometres per second and might explain certain terrestrial magnetic effects. It is hoped that further studies of hot regions and of the S component of solar radiation may throw light on the type or types of corpuscular radiation discussed by Bartels (1932) (M-regions), Waldmeier (1942) (C-regions), and Kiepenheuer (1947) (solar filaments).

VII. References

Alfvén, H. (1941).—*Ark. Mat. Astr. Fys.* **27** No. 20.
Allen, C. W. (1947).—*Mon. Not. R. Astr. Soc.* **107** : 386.
Appleton, E. V., and Hey, J. S. (1946).—*Phil. Mag.* **37** : 73.
Bailey, V. A. (1949).—*Phys. Rev.* **75** : 1104.

Bartels, J. (1932).—*Terr. Magn. Atmos. Elect.* **37** : 1.
Booker, H. G. (1935).—*Proc. Roy. Soc.* A **150** : 267.
Broxon, J. W. (1949).—*Phys. Rev.* **75** : 606.
Chapman, S. (1943-44).—*Mon. Not. R. Astr. Soc.* **103** : 117.
Christiansen, W. N., Yabsley, D. E., and Mills, B. Y. (1949).—*Aust. J. Sci. Res.* A **2** : 506.
Covington, A. E. (1948).—*Proc. Instn. Radio Engrs.* **36** : 454.
Covington, A. E. (1947).—*Nature* **159** : 405.
Covington, A. E. (1949).—*Proc. Instn. Radio Engrs.* **37** : 407.
Denisse, J. F. (1947).—*Ann. Astrophys.* **10** : 1.
Denisse, J. F. (1949).—*C.R. Acad. Sci. Paris,* **228** : 1571.
Giovanelli, R. G. (1947).—*Mon. Not. R. Astr. Soc.* **107** : 338.
Giovanelli, R. G. (1948).—*Nature* **161** : 133.
Haeff, A. V. (1949).—*Phys. Rev.* **75** : 1546.
Hagen, J. P., Jackson, R. B., McEwan, R. J., and Strang, C. B. (1948).—U.S. Naval Res. Lab. Repr., March 1948.
Hey, J. S., Parsons, S. J., and Phillips, J. W. (1948).—*Mon. Not. R. Astr. Soc.* **108** : 354.
Kiepenheuer, K. O. (1946).—*Nature* **158** : 340.
Kiepenheuer, K. O. (1947).—*Astrophys. J.* **105** : 408.
Lehany, F. J., and Yabsley, D. E. (1949).—*Aust. J. Sci. Res.* A **2** : 48.
Martyn, D. F. (1947).—*Nature* **159** : 26.
Martyn, D. F. (1948).—*Proc. Roy. Soc.* A **193** : 44.
Minnett, H. C., and Labrum, N. R. (1950).—*Aust. J. Sci. Res.* A **3** : 60.
Mount Wilson Staff (1947).—*Publ. Astr. Soc. Pacif.* **59** : 200.
Pawsey, J. L., and Yabsley, D. E. (1949).—*Aust. J. Sci. Res.* A **2** : 198.
Pawsey, J. L. (1950).—*J. Instn. Elect. Engrs.* **97** : 290.
Payne-Scott, Ruby (1949).—*Aust. J. Sci. Res.* A **2** : 214.
Payne-Scott, Ruby, Yabsley, D. E., and Bolton, J. G. (1947).—*Nature* **160** : 256.
Piddington, J. H., and Hindman, J. V. (1949).—*Aust. J. Sci. Res.* A **2** : 524.
Piddington, J. H. (1950).—*Proc. Roy. Soc.* A **203** : 417.
Piddington, J. H., and Minnett, H. C. (1949).—*Aust. J. Sci. Res.* A **2** : 539.
Ratcliffe, J. A. (1933).—*Wireless Engr.* **10** : 354.
Ryle, M. (1948).—*Proc. Roy. Soc.* A **195** : 82.
Saha, M. N., Banerjea, B. K., and Guha, V. C. (1947).—*Indian J. Phys.* **21** : 199.
Swann, W. F. G. (1933).—*Phys. Rev.* **43** : 217.
Waldmeier, M. (1942).—*Z. Astrophys.* **21** : 275.
Waldmeier, M. (1947).—*Astr. Mitt. Zurich* No. 149.

Appendix I

The Spectrum of the S Component at Metre Wavelengths

Let the range of variability of the "relatively constant component" given by Pawsey and Yabsley (1949) in Table 1 be taken to represent the upper limit of fluctuations in T_S. Let the value of T_B for all frequencies be taken as 10^6 °K. The fluctuations for 50 per cent. of the observations at 80 Mc/s. were equal to 26 per cent. of T_B (or $2 \cdot 6 \times 10^5$ °K.). The corresponding range of values of A was 2460 units so that $\alpha = 108$ °K. per unit of area.

A similar analysis by the same authors has been made for 200 Mc/s. The early data obtained in 1945–46 are rejected because they are less accurate. Data collected between March and May 1947 show a fluctuation of $(T_S + T_B)$ of 32 per cent. of T_B for 50 per cent. of the observations, and a corresponding fluctuation in A of 2640 units which, if T_B is taken as 10^6 °K., gives

$$\alpha = 120 \text{ °K. per unit of area.}$$

Similar results for August-November 1947 give a value of 87, the mean of the two being

$$\alpha = 103\ {}^\circ\text{K. per unit of area.}$$

If later results quoted by Pawsey and Yabsley are used the result is not altered significantly.

APPENDIX II

Absorption of the Extraordinary Wave

Let the direction of propagation make an angle φ with the magnetic field. Booker (1935) has shown that the propagation is quasi-longitudinal if $\sin^4\varphi/4\ \cos^2\varphi \ll [(1-x)^2+z^2]/y^2$. For $N=2\times10^9$ cm.$^{-3}$, $x=0{\cdot}4$, and $z=10^{-9}$ at the lowest frequency of 600 Mc/s. ; when $y \sim 1$ the condition is true for values of φ between zero and about 45° and between 180° and about 135°.

For quasi-longitudinal propagation, the magneto-ionic equations for the $y=1-x$ extraordinary wave are

$$\left.\begin{aligned} \mu^2-\psi^2 &= 1-\frac{x(1-y)}{(1-y)^2+z^2}, \\ 2\mu\psi &= \frac{xz}{(1-y)^2+z^2}, \end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (6)$$

where μ is the refractive index and 2ψ is the energy absorption in a distance $\lambda/2\pi$. Except in the region of $y=1$ the equations reduce to

$$\mu^2-\psi^2=\frac{1-y-x}{1-y}\,; \quad 2\mu\psi=\frac{xz}{(1-y)^2}.$$

Thus

and

$$\left.\begin{aligned} \mu^2 &= (1-y-x)/(1-y) \\ 2\psi &= \frac{xz}{(1-y-x)^{\frac{1}{2}}(1-y)^{3/2}} \end{aligned}\right\} \text{ for } (1-y-x)(1-y) \gg xz.$$

Since $z=\frac{\nu}{\omega}=\frac{\nu}{c}\cdot\frac{\lambda}{2\pi}$, the energy absorption per unit length $\varkappa$ is

$$\varkappa=\frac{\nu}{c}\cdot\frac{x}{(1-y-x)^{\frac{1}{2}}(1-y)^{3/2}} \text{ for } (1-y-x)(1-y) \gg xz.$$

Thus the equation holds except for values of y extremely close to unity and $1-x$. It is easily shown from equation (6) that the values of $\varkappa$ are then $(2cx/\nu)^{\frac{1}{2}}$ and $(2\nu x/c)^{\frac{1}{2}}$ respectively.

OBSERVATIONS OF THE SOURCE OF RADIO-FREQUENCY RADIATION IN THE CONSTELLATION OF CYGNUS

By B. Y. Mills* and A. B. Thomas*

[*Manuscript received December 29, 1950*]

Summary

The position and some of the properties of the source of radio-frequency radiation in Cygnus have been determined. The observed fluctuations in intensity have been found to correlate with activity in the F region of the ionosphere. There is no evidence that the emission from the source varies.

I. Introduction

Since its discovery by Hey, Parsons, and Phillips (1946), the discrete source of radio-frequency radiation in the constellation of Cygnus has been studied intensively by a number of observers (Bolton and Stanley 1948*a*, 1948*b* ; Ryle and Smith 1948 ; Little and Lovell 1950 ; Smith 1950 ; Stanley and Slee 1950). It is now generally agreed that the major cause of the fluctuations in intensity which it undergoes is terrestrial in origin, a conclusion which the present observations support. The actual position of the source, however, still remains to be established. Two attempts to determine the position accurately have been reported (Bolton and Stanley 1948*a* ; Stanley and Slee 1950) but the discrepancy between the determinations is greater than one degree. This apparent lack of agreement could be due either to an actual change in position in the interval between the sets of observations or to experimental errors. The latter explanation would appear to be the correct one as no observer has yet reported any systematic movement of the source. It is therefore important that further independent determinations should be carried out, preferably under differing conditions so that the errors are different.

The present series of observations has been taken on a frequency of 97 Mc/s. over a period of eight months from May 1949 to December 1949. Particular attention has been paid to sources of error, and corrections have been worked out wherever possible. The observations are being repeated at a different site with different aerials and also it is hoped to make some measurements at a much higher frequency. These observations will be described in future papers.

In addition to the position determination, the intensity of the source has been measured and upper limits of angular size, polarization, and annual parallax have been determined. Short-period variations in the apparent position have been observed. An analysis of these variations and also the amplitude fluctuations indicates that both these phenomena can be explained in terms of atmospheric effects, though other causes cannot be excluded.

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

In view of the separate origin of the fluctuations, the paper is divided into two parts, the first restricted to the properties of the source itself, and the second dealing with the fluctuation phenomena.

II. The Source

(a) *Equipment*

Most of the measurements were made with an interferometer similar in principle to that described by Ryle and Vonberg (1946, 1948). As most of the details of the interferometer system will be described elsewhere (Little and Payne-Scott—in preparation), only a brief account is given here ; a block diagram is shown in Figure 1.

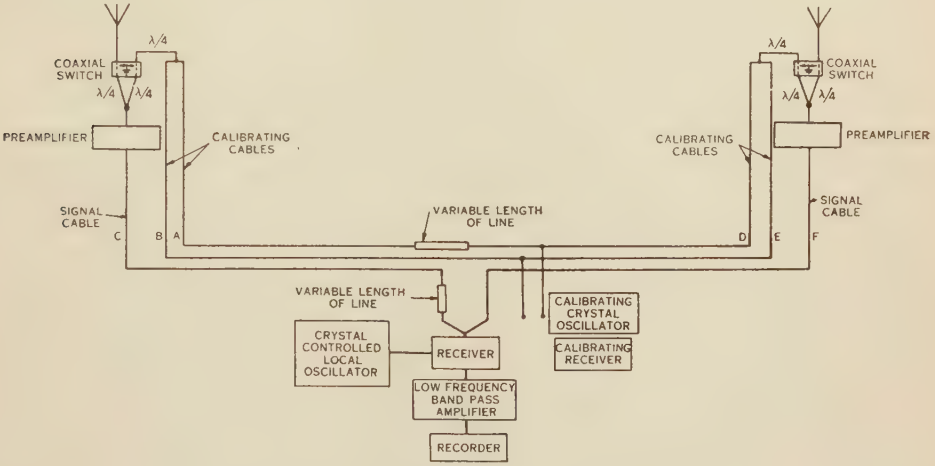

COAXIAL SWITCH
λ/4
λ/4
λ/4
PREAMPLIFIER
CALIBRATING CABLES
SIGNAL CABLE
C
B
A
VARIABLE LENGTH OF LINE
D
E
F
λ/4
COAXIAL SWITCH
λ/4
λ/4
PREAMPLIFIER
CALIBRATING CABLES
SIGNAL CABLE
VARIABLE LENGTH OF LINE
CRYSTAL CONTROLLED LOCAL OSCILLATOR
RECEIVER
CALIBRATING CRYSTAL OSCILLATOR
CALIBRATING RECEIVER
LOW FREQUENCY BAND PASS AMPLIFIER
RECORDER

Fig. 1.—A block diagram of the equipment.

Two Yagi aerials are mounted approximately 300 metres apart in an E.-W. direction and very nearly at the same elevation (difference 25 cm.). Each aerial is connected through a coaxial switch, which is used for calibration purposes as described below, to a preamplifier mounted close to the aerial to reduce feeder losses. The output from each preamplifier is fed along a coaxial cable to a centrally-situated hut, where the two outputs are combined and the resultant amplified by a 97 Mc/s. superheterodyne receiver. The bandwidth of the receiver is approximately 100 kc/s. and the local oscillator is crystal-controlled, so that the effective operating frequency is known very accurately. The rectified output of the receiver is fed to a low-frequency band-pass amplifier, which amplifies at the frequencies of the interference pattern produced by the passages of sources in front of the aerial system, i.e. frequencies of the order of 20 c/hr. The output of the amplifier is displayed on a pen recorder, together with timing marks. Part of a typical record is shown in Figure 2.

The sensitivity of the equipment is determined by the random fluctuations of output voltage. It is therefore essential that the entire amplifying equipment should be as stable as possible over the period defined by the low-frequency amplifier characteristic, in this case about 10 minutes. Voltage-regulated power supplies were used throughout, and it was found that, when the aerials were replaced by resistors of the same value, the output fluctuations were usually very much less. Thus it was the aerial fluctuation level which usually determined the sensitivity of the equipment.

(*b*) *Position Finding*

The position of a source is found in terms of its Right Ascension (R.A.) and its declination. The R.A. is determined from the time of transit, and the declination from the frequency of the interference pattern.

For R.A. measurements it is necessary to know the difference in the electrical lengths of the two paths taken by the signal from the aerials to the mixing point in the receiver hut. The method of calibration is described in detail by Little and Payne-Scott (in preparation). Briefly, the method requires the use of two pairs of coaxial " calibrating " cables (see Fig. 1). By means of switches the preamplifiers may be connected to the junction of these cables instead of to the aerials. The calibration procedure consists of comparing the electrical lengths in the east and west systems, firstly of the calibrating cables $A+B$ with the corresponding cables $D+E$; then of a calibrating cable and a signal cable (including a preamplifier) $A+C$, with the corresponding cables $D+F$; finally of the cables $B+C$, with the corresponding $E+F$. From the three relations so obtained, the difference in equivalent electrical lengths from each coaxial switch to the mixing point at the receiver may be calculated.

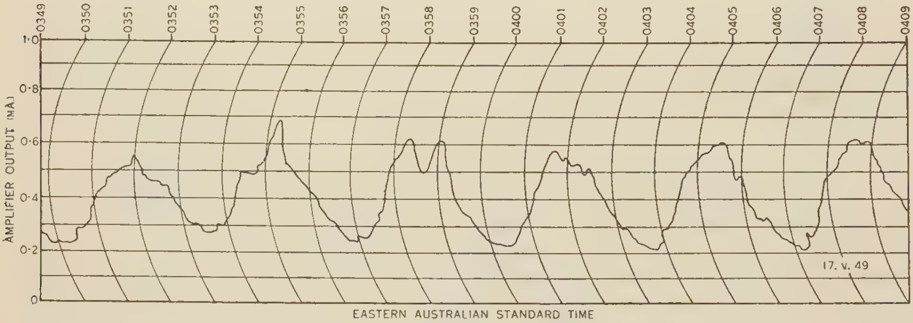


Fig. 2.—A typical record.

When so calibrated the equipment gives a series of possible times of transit of a source, each time related to the time of a minimum in the interference pattern. It must then be decided which of the positions corresponding to these times is the correct one. Had the aerials been sufficiently directive in the E.-W. direction this would have been a simple matter. However, the beam width was too wide and there was some doubt as to which of several positions was correct. It finally proved necessary to make some measurements with a different aerial spacing to resolve the ambiguities. Actually, it was not sufficiently accurate to measure the time of one minimum alone, and instead the times of five minima were averaged to give a more precise value to the central one.

When the R.A. has been found in this way, the declination is found by noting the hour angles at which minima occur. Usually the time of transit of a source does not coincide exactly with the time of a minimum and so, to make calculation easier, the minimum which occurs closest to the time of transit is taken as the central one, and hour angles are measured from the time of its occurrence. Declination values are then given by the formula

$$\cos\,\delta=\frac{n\lambda}{d\,\sin\,H_n}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

where λ is the wavelength,

d is the distance between the aerials,

n is the number of a minimum counting from the assumed central one,

and H_n is the corrected hour angle at which the nth minimum occurs.

Usually some twelve minima on either side of the central one are used to give an average value of the declination for a single observing period.

There are many possible sources of error, and appropriate corrections must be applied before the true position can be obtained.

Errors in the R.A. measurement include some due to the electrical system and some due to the site. These errors are analogous to the level, azimuth, and collimation errors of the usual transit telescope. No correction, however, is needed for aberration.

Errors in the electrical system result in collimation errors and can occur (i) in the calibrating system, due to mismatches at various points, (ii) in the aerials and aerial cables because of slightly unequal electrical properties, or (iii) in the low-frequency band-pass amplifier, where phase-shifts will alter the times of the minima. The errors due to unequal electrical properties of various components were first reduced to a minimum by careful measurement and adjustment of these components and, as a final check, the aerials and coaxial switches were separately interchanged and found to produce no detectable change in the apparent position of the source.

Siting errors due to incorrect placing of the aerials can be resolved into level and azimuth errors and the corrections calculated in the usual way (Barlow and Bryan 1944). The most serious siting error was found to be due to unequal ground-reflected rays caused by uneven terrain in the vicinity of the aerials. Because of the wide beam of the aerials and the low elevation of the source (16° at the latitude of Sydney) it is difficult to separate the direct and reflected rays. Corrections must therefore be estimated from a knowledge of the ground contours in the vicinity of each aerial and the radiation pattern of the aerial. The correction amounted to +6 seconds, but because of the difficulty of the analysis and the number of approximations necessary, there is a considerable uncertainty in this value.

In addition to these errors, random lateral deviations of the rays in the ionosphere are quite common. This effect will be discussed later.

The probable error in the R.A. measurement resulting from all these sources of error after the appropriate corrections had been applied was estimated as ±6 seconds. The errors, and measures taken to overcome them, are summarized in Table 1.

TABLE 1

ERRORS IN THE R.A. MEASUREMENT

Source of Error	Method of Correction	Comments
Differences in the electrical properties of the aerials	Careful measurement and a check by interchanging aerials	Resulting error very small
Differences between components in the calibrating system	Careful measurement and interchanging of components where practicable	Resulting error very small
Differential changes in the electrical lengths of the signal cables due to temperature changes	Calibration within a quarter of an hour of the time of transit	Resulting error negligible
Phase-shifts in the low-frequency band-pass amplifier	Calculation	Resulting error very small
Timing errors		The systematic error was less than 1 sec., but the actual reading accuracy was only ±3 sec. at best.
Aerial placement errors	Site surveyed (distances better than 2 cm., angles better than 15″) and corrections calculated	Resulting error very small
Unequal ground reflections at each aerial	Ground contours plotted and the correction estimated (+6 sec.)	The largest source of error. Estimated probable error ±5 sec.
Random variations in the times of minima, presumably due to lateral deviations in the ionosphere	Averaging a number of observations. The times of 90 minima obtained in 18 observing periods were used to obtain an average value for the R.A.	The probable error amounted to ±1½ sec.

The accuracy of the declination measurement is not as good as that of the R.A. measurement. The important factors to be considered are refraction, both as systematic and as random errors, ground reflection, aerial placement errors, uncertainty of the operating frequency, differential changes in the electrical lengths of the cables and the preamplifiers, and uncertainty in the value for the velocity of radio waves.

When vertical refraction is present the declination formula becomes

$$\cos \delta=\frac{n\lambda}{d \sin H_n}(1+R_n \tan E_n), \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

where R_n is the vertical refraction,

and E_n is the angular elevation of the source above the horizon.

Both tropospheric and ionospheric refraction must be considered. The former depends on the elevation but is independent of the frequency ; at an elevation of 15° it is approximately 5 minutes of arc (McCready, Pawsey, and Payne-Scott 1947). The latter is not known accurately but it may be calculated by assuming reasonable properties of the ionosphere (Bailey 1948). Some doubt has been cast on the value of the refraction calculated in this way at angles less than about 8° (Payne-Scott and McCready 1948), but the rather scanty experimental evidence available would suggest that at higher angles the calculated refraction is not very much in error. At a frequency of 97 Mc/s. and an elevation of 15° its value is approximately 3 minutes of arc. Adding these terms, the total refraction in the case of Cygnus's source is 8 minutes of arc. On applying this result to the declination formula the resultant correction in declination is found to be -3 minutes of arc.

The refraction is not constant, however, so that there will be a scatter in the results. In addition small lateral deviations can occur due to non-uniformity in the ionosphere. These deviations, by altering the value for H_n, can cause a disproportionately large scatter in the measured declination. The only remedy here is to take large numbers of observations.

Ground reflection effects occur as in the R.A. measurement. The correction is estimated from a knowledge of ground contours and the aerial pattern, as before. It amounts to -7 minutes of arc.

The bandwidth of the receiver results in some uncertainty in the value of the wavelength to be applied in the declination formula. If there is any asymmetry in the response curve the effective wavelength will differ from that of the maximum response. To reduce the error it is necessary to limit the receiver bandwidth ; a value of 100 kc/s. was chosen as a compromise between accuracy and sensitivity.

Differential drifts in the cable lengths, due mainly to temperature variations, were found to be serious only when observing near sunrise or sunset. Such errors can be eliminated by calibrating at intervals of about half an hour. Variations in the preamplifier phase-shifts were found to be negligible.

The value used for the velocity of radio waves is important because it affects the wavelength. Essen's (1950) value of $2\cdot99792\times10^{10}$ cm./sec. was corrected for the prevailing atmospheric conditions and the result was $2\cdot9969\times10^{10}$ cm./sec. The use of $3\cdot00\times10^{10}$ cm./sec. would produce an error of some 3 minutes of arc in the declination of Cygnus.

The probable error in the declination measurement resulting from all these sources of error after the appropriate corrections had been applied was estimated as ±3 minutes of arc. The errors, and the measures taken to overcome them, are summarized in Table 2.

TABLE 2

ERRORS IN THE DECLINATION MEASUREMENT

Source of Error	Method of Correction	Comments
Uncertainty of operating frequency due to asymmetrical response of receiver	Restriction of I.F. bandwidth to 100 kc/s.	Resulting error very small
Velocity of radio waves	Free space velocity corrected for dry air and water vapour	Resulting error negligible
Differential changes in electrical lengths of the signal cables	Calibration at half-hourly intervals	Resulting error is negligible except when observing near sunrise or sunset when rapid temperature changes can cause a small error
Aerial placement errors	Site surveyed (distances better than 2 cm., angles better than 15″) and corrections calculated	Resulting error negligible
Unequal ground reflections at each aerial	Ground contours plotted and the correction estimated (—7′ of arc)	The most serious source of error. Estimated probable error ±2′ of arc
Refraction (systematic)	Correction was calculated (—3′ of arc)	Resulting error probably very small
Refraction (random)	Averaging a large number of observations. Altogether the times of some 1000 minima obtained in 40 observing periods were used	Probable error ±1′ of arc

(c) *Characteristics of the Source*

Observations on the source included intensity, size, position, and polarization measurements. The information obtained is summarized in Table 3.

No attempt was made to measure the intensity very accurately, as the ground-reflected ray can introduce large errors which are difficult to allow for. It is not known with certainty whether the intensity of the source is constant, but the observations show that the large short-period fluctuations which are observed almost certainly originate in the ionosphere. This is discussed in detail in Section III.

On a number of occasions attempts were made to measure polarization of the radiation. Following the method of Ryle (1946, 1948), the planes of the two aerials were set at right angles. No interference pattern is then produced if the radiation is randomly polarized or—which is highly improbable—if it is plane polarized with the plane of polarization in one of the aerial planes. No interference was detected and it is concluded that the radiation is less than 3 per cent. polarized. This is in agreement with Ryle who also could detect no polarized component.

TABLE 3

OBSERVATIONS ON CHARACTERISTICS OF THE SOURCE

Property Observed	Value	Estimated Probable Error
Intensity	$1 \cdot 5 \times 10^{-22}$ W./m.$^{-2}$/c/s.	$\pm 3 \times 10^{-23}$ W./m.$^{-2}$/c/s.
Angular size	Less than 10 min. of arc	
Position (Epoch 1950)	R.A. 19 hr. 57 min. 37 sec. Declination 40° 34′	±6 sec. ±3′
Polarization of the radiation	Less than 3% polarized	
Annual parallax	Less than 1 min. of arc	

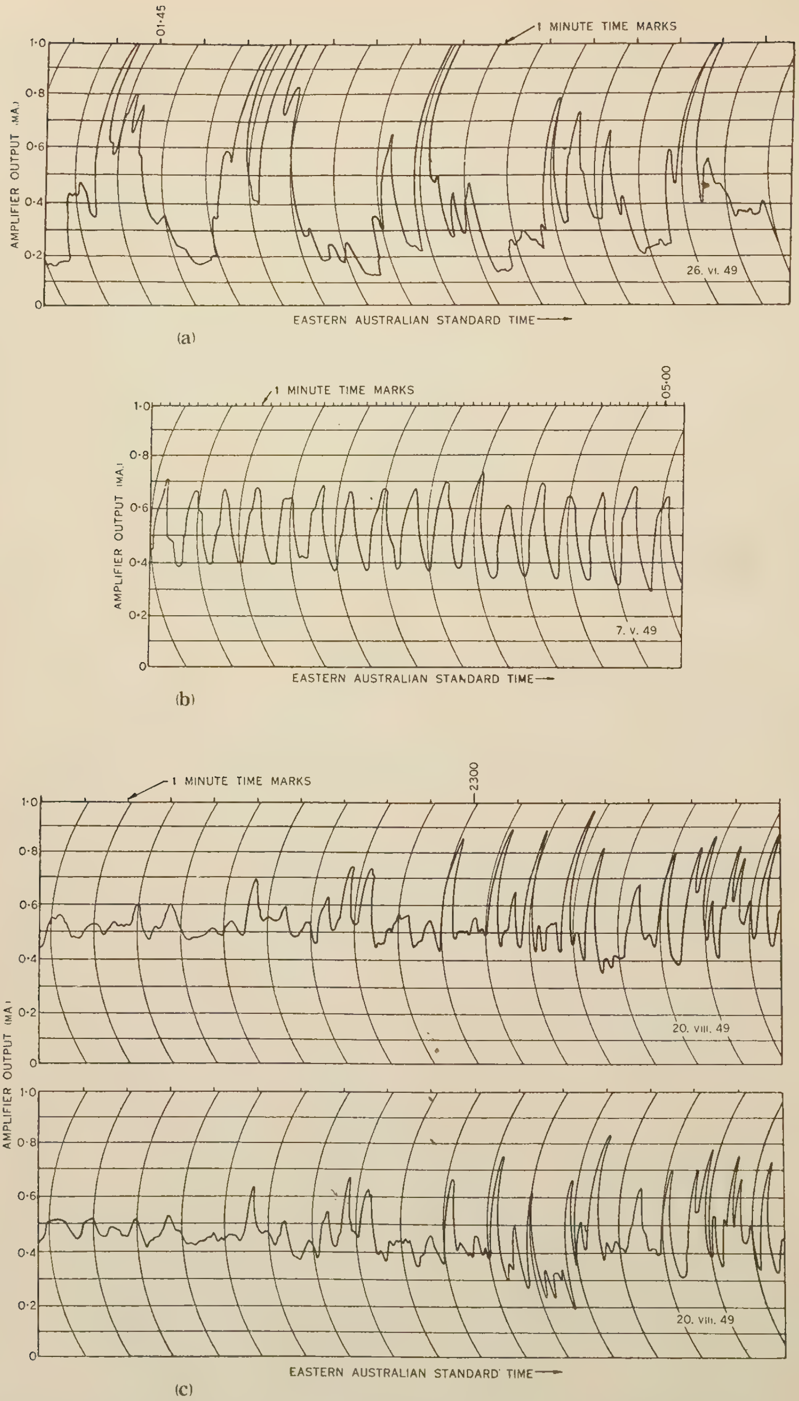
The maximum angular size was estimated only very approximately by measuring the ratio between intensity fluctuations on the maxima and minima of the interference pattern. Later it is shown that a consideration of fluctuation phenomena suggests an angular diameter less than 3 minutes of arc.

Within the probable position field of the source (see Plates 1, 1A), there is no prominent object which could be expected to produce the observed radiation. There are a number of faint nebulae, one in particular, of about the 15th magnitude, being within the probable error limits of the source at a position of

R.A. 19 hr. 57 min. 41 sec.
Dec. 40° 35′.

Minkowski, of Mount Wilson and Palomar Observatories, in a private communication, has identified this nebula as extra-galactic, at an estimated distance of 10^7 parsec. It seems unlikely that it could be the source of the radiation, as the effective temperature required would be at least 10^{10} °K. and it is unlikely that so large an object should be so abnormal, particularly in view of the similar appearance of the other nebulae near by. The conclusion to be drawn is rather that there is little interstellar matter in the vicinity, and hence it is unlikely that the source is located in a nearby bright star which is dimmed by interstellar absorption.

No evidence of annual parallax could be detected ; the limit of detection being about 1 minute of arc. This means that the distance of the source from the Earth is at least 40 times the radius of Pluto's orbit. It is therefore concluded that the source is probably unconnected with the solar system. This contradicts



(a)

(b)

(c)

Fig. 3.—(*a*) An example of "fast" fluctuations. (*b*) An example of "slow" fluctuations. (*c*) Examples of simultaneous records on separate receivers spaced 300 metres apart.

the suggestion of Menzel and Crowley (1950) that the discrete sources of radio-frequency energy are nearby comets, with parallaxes of the order of 1°.

On the evidence presented it would seem most likely that the source is located in some nearby faint star of abnormal properties. A proper-motion search of the field would therefore be of the greatest interest.

III. The Fluctuations

(*a*) *Observations*

One of the most interesting properties of the source in Cygnus is the large fluctuations in amplitude which sometimes occur and which led to its discovery (Hey, Phillips, and Parsons 1946). It is now fairly well established that the fluctuations are terrestrial and probably ionospheric in origin (Little and Lovell 1950 ; Smith 1950 ; Stanley and Slee 1950), although no direct correlation with ionospheric data has previously been found possible. The present series of observations does, however, show a significant correlation between the fluctuations and short-period activity in the *F* region.

In addition to the fluctuations, deviations in the apparent position of the source are quite common, that is the lobe minima do not always occur at the expected times. Times of individual minima have on occasions been more than a minute in error. It is shown later that these deviations are related to the fluctuations and presumably have the same origin.

Some typical records of fluctuation phenomena are shown in Figure 3. It is convenient to classify the fluctuations as " fast " or " slow " depending on whether their duration is shorter or longer than the lobe period. The former show one or more minima within a lobe, the latter show variations in intensity between successive lobes. In Figure 3 (*a*) we have an example of very intense " fast " fluctuations, which had an average period of about half a minute. In Figure 3 (*b*) a record showing " slow " fluctuations has been redrawn with a different time scale to make the effect more obvious. The slow fluctuations are not comparable in magnitude with the fast fluctuations of Figure 3 (*a*).

In Figure 3 (*c*) are shown two records of intense fluctuations on single receivers spaced 300 m. apart. A few such observations were made to examine the spatial distribution of the amplitude fluctuations by connecting separate receivers to each of the interferometer aerials. It can be seen that the records are not identical although they are very similar. These records were taken on a night when the fluctuations were very intense. In addition some measurements were made in conjunction with Mr. J. G. Bolton of this Laboratory, in which the separation between receivers was 30 km., but in this case there was no detailed correlation whatever between the fast fluctuations.

(*b*) *Analysis of the Observations*

To study various possible correlations with geophysical data three indices were computed—a " fast " fluctuation index, a " slow " fluctuation index, and an index of apparent position deviation. Each is based on a single observation period of an hour or two on either side of culmination of the source. The " fast " fluctuation index, which is similar to that used by Stanley and Slee

(1950), is obtained by measuring the maximum and minimum intensities within each lobe and calculating the ratio of the difference between these intensities to their mean value. The separate ratios for each lobe are then averaged to obtain the fluctuation index of a single observation period. The index may thus have a value between 0 and 2 ; the average of all the observations was approximately 0·4. The " slow " fluctuation index is a measure of the scatter of the average intensity of individual lobes from the mean intensity of all the lobes during the observation period, and is obtained by calculating the probable error of the mean intensity. The " deviation " index is a measure of the scatter in lobe-minima times over a period of about one hour and is obtained by calculating the probable error in declination of a transit observation.

Correlations of these indices were attempted with magnetic data, the occurrence of sporadic E ionization, and the critical frequency of the F region, without success. A direct correlation was obtained, however, between the fast fluctuation index and F-region activity as measured by an index supplied by Mr. G. H. Munro of the Radio Research Board. This index, which takes the values 0, 1, 2, 3, was based on observations at a fixed frequency of 5·8 Mc/s. The prevalence of activity such as " Z " 's (Munro 1950), diffuse F-region echoes, and fluctuations in intensity and height of echoes, was estimated for a given day and an index given for the day. This procedure suffers somewhat from the fact that the ionospheric observations did not usually cover the same period as the Cygnus observations. However, with nearly sixty observations, the correlation between the indices was 0·36 ; the probability that such a coefficient could arise by chance is less than 0·01 (Fisher 1938). No regular seasonal effect could be detected, so that it is considered that there is a real physical relation between the two phenomena.

In addition the fluctuations were studied alone to see if any deductions could be made from their properties. Correlations between the various indices were as follows :

$$r_{12}=0\cdot45,\ r_{13}=0\cdot57,\ r_{23}=0\cdot71,$$

where the fast fluctuation index is designated by 1, the slow fluctuation index by 2, and the deviation index by 3. The significant correlations between all three pairs of indices point to a common origin.

In order to test the hypothesis that the fluctuations are the result of ray deviations in the intervening medium, partial correlations were calculated (Fisher 1938). The partial correlations in effect give the correlation between two of the three related phenomena when the third is held constant. The results were as follows :

$$r_{12\cdot3}=0\cdot08,\ r_{13\cdot2}=0\cdot35,\ r_{23\cdot1}=0\cdot62.$$

These results indicate that the causal link between the fast and slow fluctuations occurs only through the deviation index, for when this is held constant the correlation ($r_{12\cdot3}$) is negligible. This is consistent with the above hypothesis. The fact that $r_{13\cdot2}$ is lower than $r_{23\cdot1}$ may not be significant, but it is reasonable because position measurements are made only at intervals of 3½ minutes and in these circumstances slow deviations are more likely to affect the deviation index than fast deviations.

A correlation of 0·6 was also found between the fast fluctuation index and the average period of the fast fluctuations.

(*c*) *Discussion of the Observations*

After an examination of these results a number of inferences can be made concerning the mechanism involved in the production of fluctuations. The correlation between fluctuation index and ionospheric-activity index suggests strongly that the fluctuations have their origin in the *F* region. The obvious explanation, and one which is supported by the partial correlations, is that they are the result of irregularities of refractive index in the *F* layer. Irregularities may also be deduced from the occurrence of diffuse *F* echoes.

Assuming such irregularities exist we may attempt to obtain some idea of the order of their size as follows. When the fluctuations are intense their minimum period is of the order of 20 seconds. If it is assumed that at this time there is a wind velocity of 10 km./min. (Munro 1950) the minimum size of the irregularities will be the product of these quantities, approximately 3 km. Similarly a maximum size may be estimated, since the longest periods observed when very fast fluctuations are present is about 6–10 minutes (that is two or three lobes), corresponding to a maximum size of about 50–100 km. Longer periods (up to 15–20 minutes) which have been observed when there is little " fast " activity could well be due to lower wind velocities.

The minimum size of the irregularities may also be deduced from the spaced receiver measurements. In these a separation of 0·3 km. between receivers gave good, but not perfect, correlation between fast fluctuation records, as shown in Figure 3 (*c*). On the other hand, some measurements made when the spacing was nearly 30 km. gave no correlation at all for the fast fluctuations. Thus we may say that the minimum size is much greater than 0·3 km. and much less than 30 km. Little and Lovell (1950), observing in England, have quoted correlations of between 0·5 and 0·95 with a separation of 3·9 km. Their results would therefore indicate an average size which is somewhat larger than 4 km. This is consistent with the above conclusions.

This analysis ignores the effects of diffraction and interference, which actually are likely to be very important in the production of intense short-period fluctuations. The first Fresnel zone at the *F* region and at an elevation of 16° has a diameter of approximately 3 km., so that the minimum irregularity size determined above is probably not a real minimum. However, the fact that such a short period as 20 seconds is comparatively rare shows that the irregularities are usually larger than this. The fluctuations are therefore thought of as arising mainly from the large scale irregularities acting as a random lens system, which causes convergence or divergence of the rays from the source.

Because of the correlation between the fluctuation intensity and the average period of the fluctuations, a turbulence mechanism would seem likely as the cause of the irregularities. It is therefore of interest to examine the possibility of a turbulence which has eddies of about the same size as the irregularities discussed above, being set up by the presence of the high-velocity winds which occur at this height. Let us assume an eddy size of 3 km. and calculate its

Reynolds number. Assuming a height of 300 km. and a wind velocity of 10 km./min., and obtaining the density and viscosity from published data (Mitra 1947), we find that the Reynolds number is of the order of unity. The production of turbulence normally requires a Reynolds number of 1000 and the Reynolds number of the smallest eddies in a free turbulence is unlikely to be smaller than one order of magnitude less than this value. Thus to explain a turbulence with such a small scale, some mechanism must be found by which energy may be supplied directly to the small eddies. Possibly such a mechanism may be found in the rapid change of pressure with height, and the presence of a magnetic field which tends to give a vertical component to any horizontal motion of the ions. Failing some explanation along these lines, the idea of turbulence, however attractive as a solution, must be abandoned.

(*d*) *Source Size*

With the above model of an irregular F layer we may make some deductions concerning the size of a source from a study of its fluctuations. Clearly, irregularities which subtend a much smaller angle at the receiver than does the source will not affect the intensity of the source. Thus sources of a large angular size will show only slow and relatively small fluctuations, while small sources will sometimes show fast and intense fluctuations in addition. Since the angle subtended by the irregularities is a function of the elevation of the source above the horizon, it may be expected that the shortest fluctuation period of a given source will also show some dependence on elevation. Thus it is quite possible that a source will show only slow fluctuations when rising but fast and slow fluctuations when culminating.

To put the matter on a more quantitative basis let us calculate the angle subtended by an irregularity at zero angular elevation, and at a height of 300 km. above the Earth's curved surface. For a 5-km. irregularity this is about 3 minutes of arc. The minimum period of the Cygnus fluctuations would actually appear to be shorter when rising than when culminating, suggesting that the source is much smaller than 3 minutes of arc in angular size and that, as one might expect, diffraction and interference effects are of some importance at low elevations. Some other sources which have been observed do show effects which suggest an angular size comparable with or greater than 3 minutes of arc; these observations will be discussed in a later paper.

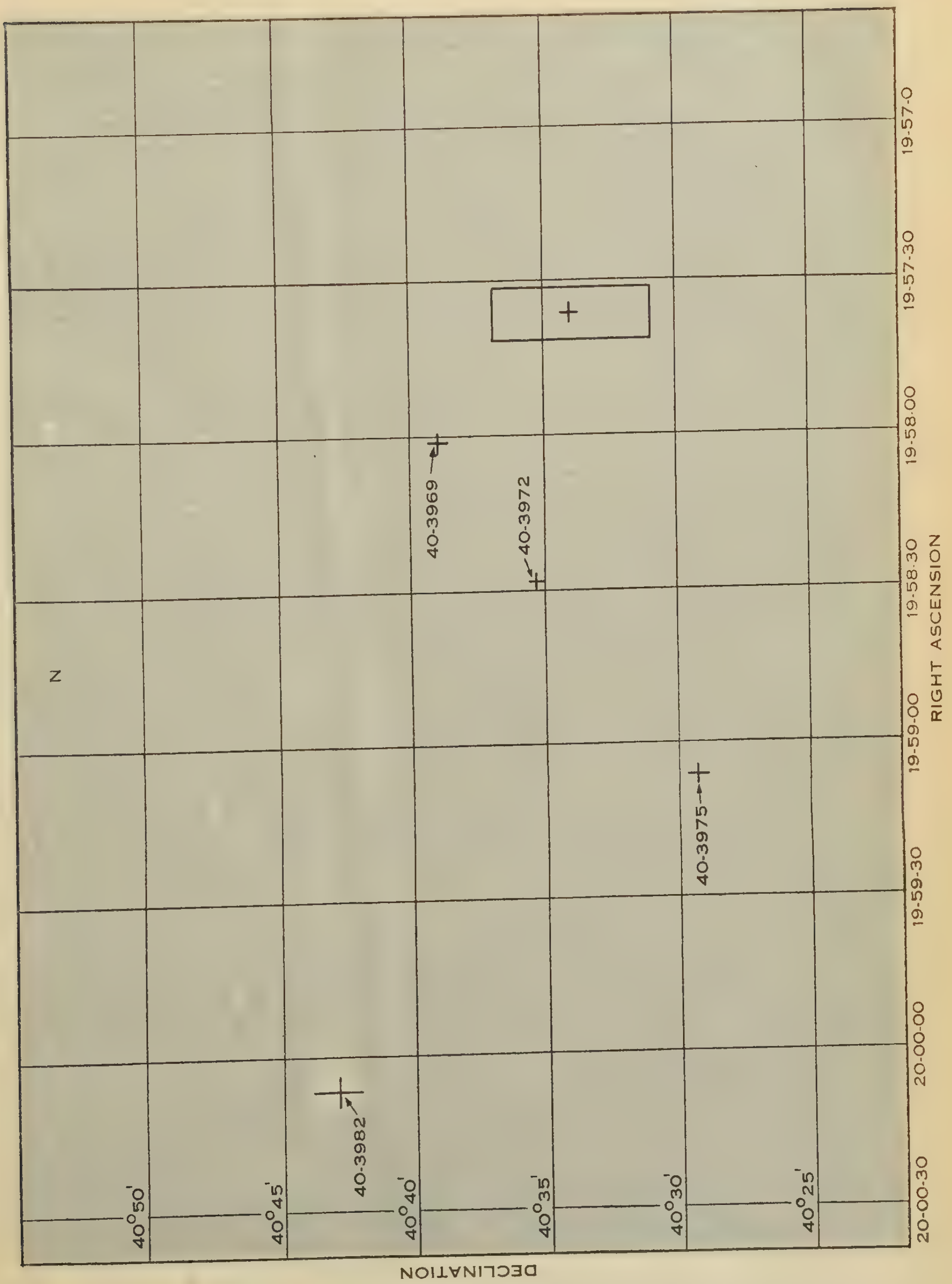
IV. CONCLUSIONS

This series of observations indicates that the discrete source in Cygnus has a position given by

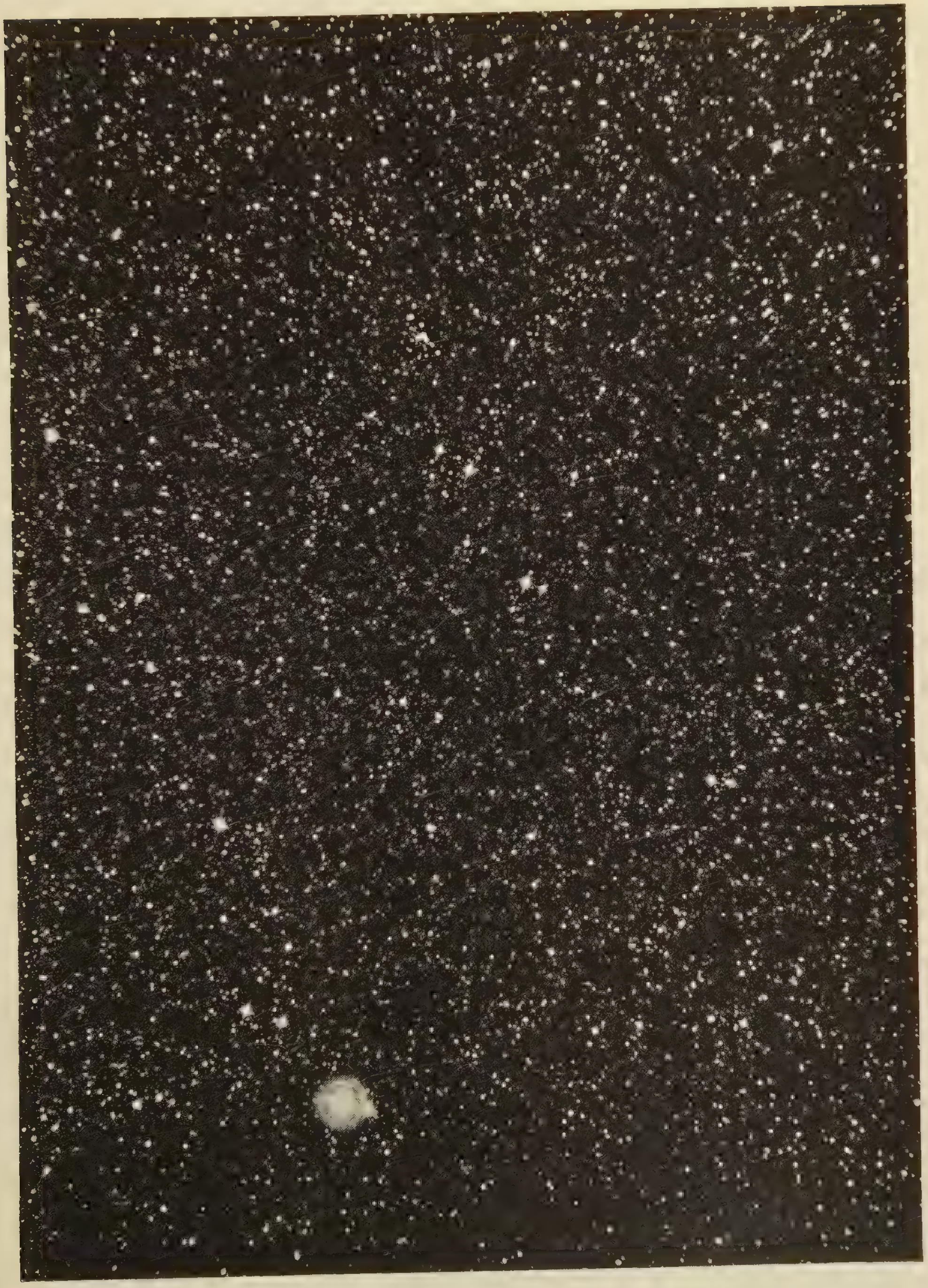
R.A. 19 hr. 57 min. 37 sec.
Dec. 40° 34′

with a probable error of 6 seconds in R.A. and 3 minutes of arc in declination. A rectangle indicating this position and the probable error limits is shown in Plate 1. Within this region there is no prominent or unusual celestial object with which the source could be identified. The presence of a number of extragalactic nebulae in the vicinity indicates the absence of any obscuring interstellar

SOURCE OF RADIO-FREQUENCY RADIATION IN CYGNUS


19·57·0
19·57·30
19·58·00
19·58·30
19·59·00
19·59·30
20·00·00
20·00·30
RIGHT ASCENSION
40·3969
40·3972
40·3975
40·3982
N
40°50'
40°45'
40°40'
40°35'
40°30'
40°25'
DECLINATION

SOURCE OF RADIO-FREQUENCY RADIATION IN CYGNUS



matter which might dim a bright star, while permitting its radio-frequency radiation to pass through unhindered. The absence of any large annual parallax indicates that the source is unlikely to be part of the solar system.

The apparent fluctuations in the intensity of the source have been shown to originate at least partly, and probably mainly, in the F region of the ionosphere, which is pictured as having fairly large-scale irregularities of the order of 5 to 100 km. across.

V. REFERENCES

BAILEY, D. K. (1948).—*Terr. Magn. Atmos. Elect.* **53** : 41.

BARLOW, C. W. C., and BRYAN, G. H. (1944).—" Elementary Mathematical Astronomy." (University Tutorial Press : London.)

BOLTON, J. G., and STANLEY, G. J. (1948).—*Aust. J. Sci. Res.* A **1** : 58, 139-48.

BOLTON, J. G., and STANLEY, G. J. (1948).—*Nature* **161** : 312.

ESSEN, L. (1950).—*Nature* **165** : 443.

FISHER, R. A. (1938).—" Statistical Methods for Research Workers." 7th Ed. pp. 198 *et seq.* (Oliver and Boyd : Edinburgh.)

HEY, J. S., PHILLIPS, J. W., and PARSONS, S. J. (1946).—*Nature* **157** : 296-7.

LITTLE, A. G., and PAYNE-SCOTT, RUBY.—" An interferometer for locating bursts of solar noise at 97 Mc/s." (in preparation).

LITTLE, C. G., and LOVELL, A. C. B. (1950).—*Nature* **165** : 422.

McCREADY, L. L., PAWSEY, J. L., and PAYNE-SCOTT, RUBY (1947).—*Proc. Roy. Soc.* A **190** : 357.

MENZEL, D. K., and CROWLEY, D. J. (1950).—*Nature* **165** : 443.

MITRA, S. K. (1947).—" The Upper Atmosphere." p. 514. (Royal Asiatic Society of Bengal.)

MUNRO, G. H. (1950).—*Proc. Roy. Soc.* A **202** : 208.

PAYNE-SCOTT, RUBY, and McCREADY, L. L. (1948).—*Terr. Magn. Atmos. Elect.* **53** : 429.

RYLE, M., and SMITH, F. G. (1948).—*Nature* **162** : 462-3.

RYLE, M., and VONBERG, D. D. (1946).—*Nature* **158** : 339.

RYLE, M., and VONBERG, D. D. (1948).—*Proc. Roy. Soc.* A **193** : 98.

SMITH, F. G. (1950).—*Nature* **165** : 422.

STANLEY, G. J., and SLEE, O. B. (1950).—*Aust. J. Sci. Res.* A **3** : 234-50.

EXPLANATION OF PLATES 1, 1A

The star field in the neighbourhood of the source. The position of the source and the probable error limits are shown by the cross and rectangle respectively. Identifying stars marked by crosses are from the Bonn Durchmusterung Catalogue. (Photograph by courtesy of the Mount Wilson and Palomar Observatories, California, U.S.A.)

THE INFLUENCE OF SLIT-WIDTH ON THE SHAPE AND INTENSITY OF INFRA-RED ABSORPTION BANDS

By J. B. Willis*

[*Manuscript received January 5, 1951*]

Summary

Making certain assumptions as to the shape of infra-red absorption bands and the shape of the slit function of the monochromator, expressions are obtained for the dependence on spectrometer slit-width of the intensity and half-width of absorption bands. Experimental data to confirm the accuracy of these deductions are presented.

I. Introduction

In a recent paper (Walsh and Willis 1950) from this Laboratory, the influence of temperature on the breadths of the infra-red absorption bands of a number of substances in the solid and liquid states was discussed. It was concluded that the narrowing of the bands at low temperatures claimed by previous workers (Avery and Ellis 1942 ; Avery and Morrison 1947) was not of general occurrence. In connection with this work calculations were carried out to determine how the observed band-width varies with the slit-width of the spectrometer. If, owing to the finite slit-width of the spectrometer, the observed band-width is considerably greater than the true width, then any variation of the true width with temperature will be partially masked. As there is considerable interest in the effect of low temperatures on infra-red absorption spectra, it was thought that these calculations might be of value in suggesting (i) a new method of approaching the problem of slit-width errors in spectrophotometry, and (ii) the order of magnitude of these errors for some typical absorption bands using a well-known commercial infra-red spectrometer.

Previous work on slit-width errors has been largely concerned with spectral *lines*, such as the rotational lines in the vibration-rotation bands of simple gaseous molecules. For instance, Dennison (1928) has investigated the effect of finite slit-width on the shape of isolated absorption lines, and Elsasser (1938) the case of a sequence of equidistant, identical absorption lines.

The case of a continuous absorption band, as found in the spectrum of a solid, liquid, or solution, does not seem to have been treated until recently. Hardy and Young (1949) have made a rigorous treatment of slit-width errors for the general case of a continuous absorption curve, and have indicated the method of applying it to spectrometric problems. Their method, however, requires the use of a line source of about the same frequency as the absorption band to be measured, and so does not seem readily applicable in the infra-red region. While the present work was in progress, Eberhardt (1950) showed

---

* Division of Industrial Chemistry, C.S.I.R.O., Melbourne.

that by assuming the absorption curve to be a parabola near maximum absorption, the observed transmission at that point can be accurately corrected if the slit function of the spectrometer is known. The effect of finite slit-width on the *breadth* of absorption bands was not dealt with.

The usual type of infra-red spectrometer works on the following principle : radiation from a continuous source having approximately black-body characteristics traverses the specimen, passes through the entrance slit and is dispersed by the prism. From the spectrum thus formed, a small frequency range is selected by the exit slit, which normally has the same width as the entrance slit, and is focused on the detector, whose output is amplified and recorded on a chart.

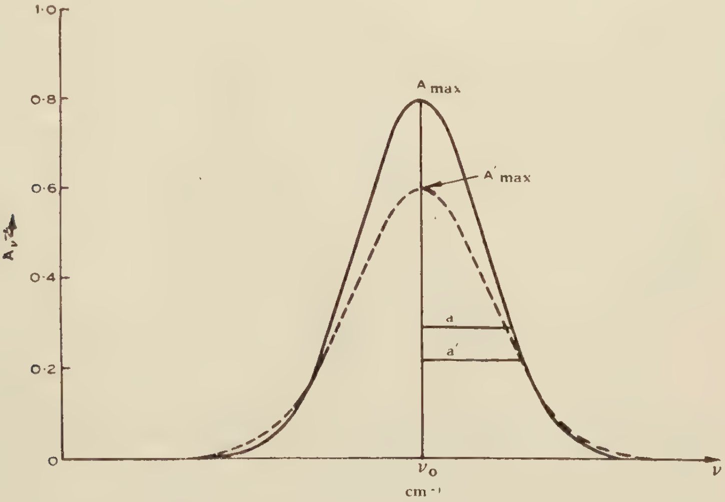


Fig. 1.—Effect of measuring a Gaussian-shaped absorption curve with a spectrometer of finite resolving power.

For a quantitative discussion of the effect of finite slit-width on the shape of an absorption band a knowledge of two things is required :

(i) The equation of the true absorption curve being studied, i.e. the curve which would be obtained with a spectrometer of infinitely high resolution.

(ii) The slit function of the monochromator, i.e. the relation between light flux and frequency in the beam of almost monochromatic light furnished by the exit slit.

Care was taken to select for study only bands which were free from overlap with adjacent bands. It is found that for solid and liquid specimens the shape of such bands, even when obtained using an instrument with wide slits, approximates closely to the familiar Gaussian error curve (see Richards and Burton 1949). This suggested that the true band shape might be such a curve, and it is with this assumption that the equations will be worked out. Figure 1 shows qualitatively the effect of measuring a Gaussian absorption curve with a spectrometer of finite resolving power, the full and dotted lines representing respectively the true and observed absorption curves.

Over the small frequency range covered by the exit slit it may be assumed, to a first approximation, that the dispersion is constant and that the light flux incident on the entrance slit is independent of frequency, i.e. that the black-body curve of the light source is flat over this range. (For the absorption bands studied in the present work the slope of the black-body curve in the neighbourhood of the band was of the order of 1 per cent. or less per cm.$^{-1}$.) Under these conditions, assuming ideal optics, the slit function will be triangular ; this is treated below as case (*a*). To compensate somewhat for optical imperfections, which will have the effect of rounding off the angles of the triangle, the slit function is taken as a Gaussian curve whose parameters are such that the peak of the curve coincides with the apex of the triangle and the area under the curve equals the area of the triangle, i.e. the total amount of energy passed by the slit in both cases is the same (see Fig. 2). This is treated as case (*b*). Clearly, other assumptions could be made as to the analytical form of the slit function, but the two forms chosen here give results in accordance with the experimental figures and have the merit that they can be handled mathematically with relative ease. It may be noted that Eberhardt (1950) used a triangular slit function and Dennison (1928) a Gaussian one, both withgood results.

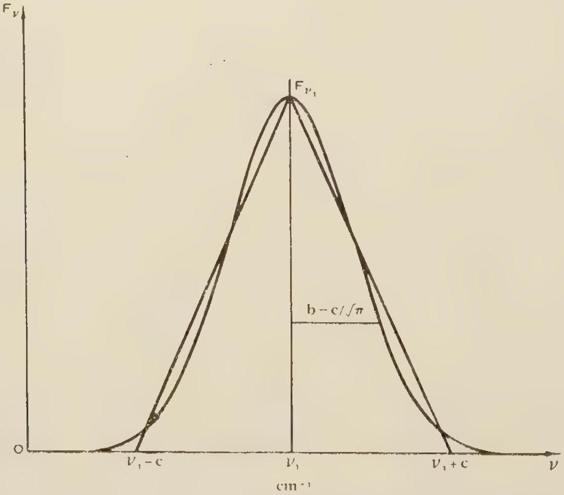


Fig. 2.—Possible forms of slit function : (*a*) Triangular, (*b*) Gaussian. The monochromator is set on a nominal frequency ν_1, and the energy passed by the exit slit is the same in both cases.

With these assumptions as to the true form of the absorption band and the form of the slit function of the monochromator, expressions are derived giving :

(i) The true half-width of the absorption band in terms of the observed half-width.

(ii) The true absorption intensity at the peak of the band in terms of the observed intensity at that point.

The spectral slit-width is defined as half the total frequency range passed by the exit slit (Barnes *et al.* 1945). It is the sum of two terms, one determined by the mechanical slit-width and one by the limiting resolving power of the prism. The latter term will vary somewhat with the mechanical slit-width, although it is normally calculated only for zero slit-width. However, in all the cases examined in the present paper the term due to mechanical slit-width (even with the narrowest slits used) is considerably greater than that due to prismatic resolution, so that small variations in the latter may be neglected.

Scattering in the instrument is neglected. The image of the exit slit is supposed to lie within the sensitive element of the detector, which is assumed to be non-selective and linear. The spectrum is supposed to be scanned sufficiently slowly for the detector, amplifier, and recorder to follow accurately the variations in the radiation flux transmitted by the specimen.

It must be understood that what is referred to in this paper, and indeed in most infra-red literature, as the " absorption curve " is defined by the *transmission* of the specimen, and so includes any contribution from reflexion and scattering. However, the term " absorption band " will be retained for convenience in the present paper.

II. Derivation of the Expressions

The equation of the true absorption curve is taken as

$$A_\nu = A_{max} \exp[-(\nu-\nu_0)^2/a^2],$$

where A_ν and A_{max} are defined as $1-T_\nu$ and $1-T_{max}$ respectively, T_ν and T_{max} being the fractions of light transmitted at any frequency ν and at the peak of the curve (i.e. at ν_0). A'_ν and A'_{max}, T'_ν and T'_{max} are the corresponding quantities for the observed absorption curve obtained with finite slits. The constant a is half the band-width at $1/e$ of A_{max}. It bears a linear relationship to the " half-width ", i.e. the band-width at $A_{max}/2$, which is

$$2a\sqrt{\log_e 2} = 1 \cdot 665a.$$

Similarly, a' is half the observed band-width at $1/e$ of A'_{max}.

For a given frequency setting ν_1 of the monochromator, the light flux F_ν is defined so that $F_\nu d\nu$ is the energy transmitted by the monochromator in the frequency range ν to $\nu+d\nu$. G_{ν_1} is the energy received by the detector. In the absence of reflexion or absorption by the specimen, $G_{\nu_1} = \int F_\nu d\nu$.

(a) *Triangular Slit Function*

With the monochromator set at a nominal frequency ν_1, and a spectral slit-width c cm.$^{-1}$ defined as above:

$$F_\nu = F_{\nu_1}\left(1 - \frac{\nu_1-\nu}{c}\right), \quad \nu_1 - c < \nu < \nu_1 ;$$

$$F_\nu = F_{\nu_1}\left(1 + \frac{\nu_1-\nu}{c}\right), \quad \nu_1 < \nu < \nu_1 + c.$$

F_ν is non-zero only for $\nu_1 - c < \nu < \nu_1 + c$.

The energy received by the detector at the monochromator setting ν_1 will be the integral of the product of the energy supplied by the slit over a range ν to $\nu+d\nu$ and the transmission of the specimen at the frequency ν.

Now

$$G_{\nu_1}=T'_{\nu_1}\int_{\nu_1-c}^{\nu_1+c}F_\nu d\nu=\int_{\nu_1-c}^{\nu_1}F_\nu(1-A_\nu)d\nu+\int_{\nu_1}^{\nu_1+c}F_\nu(1-A_\nu)d\nu.$$

Also, $\int_{\nu_1-c}^{\nu_1+c}F_\nu d\nu=F_{\nu_1}c$, and the other integrals can be evaluated in terms of error functions (see, for example, Carslaw and Jaeger 1947). Replacing T'_{ν_1} by $1-A'_{\nu_1}$ and putting $(\nu_1-\nu_0)/a=p$ and $c/a=q$ the following equation is obtained :

$$A'_{\nu_1}=\frac{A_{max}\sqrt{\pi}}{2q}\cdot Q, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

where

$$Q=\operatorname{erfc}(p-q)-\operatorname{erfc}(p+q)+\frac{p}{q}[2\operatorname{erfc} p-\operatorname{erfc}(p-q)-\operatorname{erfc}(p+q)]$$
$$+\frac{1}{q\sqrt{\pi}}[e^{-(p-q)^2}+e^{-(p+q)^2}-2e^{-p^2}].$$

Putting $\nu_1=\nu_0$, the relation between A'_{max} and A_{max} is obtained, namely

$$A'_{max}=\frac{A_{max}\sqrt{\pi}}{q}\left\{\operatorname{erf} q-\frac{1}{q\sqrt{\pi}}[1-\exp(-q^2)\right\}. \quad \ldots\ldots \quad (2)$$

To determine a by means of equation (1), A'_{ν_1} is replaced by A'_{max}/e and A'_{max} is expressed in terms of A_{max} by equation (2). In this case $(\nu_1-\nu_0)/a=a'/a$, and thus

$$2/e=Q\Bigg/\left(\operatorname{erf} q-\frac{1}{q\sqrt{\pi}}[1-\exp(-q^2)]\right). \quad \ldots\ldots \quad (3)$$

Equation (3) can be solved for a by trial and error, since c is known and a' is known from measurement of the observed absorption band. Comparison of the values of a' for various slit-widths with the calculated value of a gives a direct measure of the dependence of the observed width of the absorption band on the spectral slit-width.

(b) *Gaussian Slit Function*

If it is now assumed that the slit function has the form

$$F_\nu=F_{\nu_1}\exp[-(\nu-\nu_1)^2/b^2],$$

but that the total energy passed by the exit slit for a given setting ν_1 of the monochromator is the same as for the triangular slit function, then

$$\int_{-\infty}^{+\infty}F_{\nu_1}\exp[-(\nu-\nu_1)^2/b^2]d\nu=F_{\nu_1}c,$$

so that $b=c/\sqrt{\pi}$.

Following the procedure of case (*a*):

$$G_{\nu_1}=\int_{-\infty}^{+\infty}F_\nu(1-A_\nu)d\nu,$$

so that

$$T'_{\nu_1}=\frac{1}{F_{\nu_1}c}\int_{-\infty}^{+\infty}F_\nu(1-A_\nu)d\nu,$$

and thus

$$A'_{\nu_1}=\frac{A_{max}}{c}\int_{-\infty}^{+\infty}\exp\left\{-\left[\frac{(\nu-\nu_0)^2}{a^2}+\frac{(\nu-\nu_1)^2}{c^2}\pi\right]\right\}d\nu.$$

Evaluation of this integral gives

$$A'_{\nu_1}=\frac{A_{max}\,a\exp\left[-(\nu_1-\nu_0)^2\Big/\left(a^2+\frac{c^2}{\pi}\right)\right]}{\left(a^2+\frac{c^2}{\pi}\right)^{\frac{1}{2}}}. \qquad \ldots\ldots \quad (4)$$

Putting $\nu_1=\nu_0$, the relationship between A'_{max} and A_{max} is

$$A'_{max}=\frac{A_{max}\,a}{\left(a^2+\frac{c^2}{\pi}\right)^{\frac{1}{2}}}. \qquad \ldots\ldots\ldots\ldots \quad (5)$$

To determine *a* from equation (4), put $A'_{\nu_1}=A'_{max}/e$ and substitute for A'_{max} in terms of A_{max} by means of equation (5). As before, $\nu_1-\nu_0=a'$, so that

$$a=\left(a'^2-\frac{c^2}{\pi}\right)^{\frac{1}{2}}. \qquad \ldots\ldots\ldots\ldots \quad (6)$$

Equations (5) and (6) correspond to equations (2) and (3) for the triangular slit function.

III. Experimental Verification

To test the correctness of the formulae obtained here, typical liquid and solid absorption bands were carefully measured at different slit-widths. The bands chosen were (i) the 715 cm.$^{-1}$ band of methylene iodide, (ii) the 985 cm.$^{-1}$ band of methane sulphonyl fluoride, and (iii) the 1017 cm.$^{-1}$ band of (solid) 1-hydroxyanthraquinone. The two liquids were contained in conventional fixed-length cells with rock-salt windows, while the solid was vacuum-sublimed on to a rock-salt plate to ensure a uniform layer of small grain size.

The instrument used was a Perkin-Elmer Infra-Red Spectrometer, Model 12C. The light from the globar source was chopped at a frequency of 13 c/s. and the output from A.C. thermocouple (time-constant 0·1 sec. for 90 per cent. response) fed to an amplifier tuned to this frequency, whose output was led to the recorder.

The slits were carefully adjusted to close simultaneously and to be free from zero error, and the rock-salt prism repolished to reduce the proportion of scattered light to a minimum. A lithium fluoride shutter was used in recording the 715 cm.$^{-1}$ band, and a solid shutter for the other two bands.

The bands were scanned at slow speed to ensure full response by the recorder. The zero was perfectly constant, and the black-body curve was drawn from measurements on the black-body curve for a blank cell.

The value of $2a'$ for a given slit-width was obtained directly from the chart by measuring the frequency separation of the two points where the absorption intensity was $1/e$ of the maximum. For the treatment of case (*a*) a trial value of a was chosen, somewhat less than a', and the value of the right-hand side of equation (3) calculated, after which a was adjusted until the two sides balanced.

Tables 1, 2, and 3 show the experimental figures and the values of a and A_{max} calculated from them on the assumption of (*a*) a triangular slit function, and (*b*) a Gaussian slit function. The approximate slope of the black-body curve is indicated above each table as the percentage change per cm.$^{-1}$ of incident radiation energy at the frequency of the band-peak. The values of a', a, and c have been rounded off to the nearest 0·05 cm.$^{-1}$ and the values of A'_{max} and A_{max} to the nearest 0·005, as they are limited by the accuracy of measurement.

TABLE 1

715 CM.$^{-1}$ BAND OF METHYLENE IODIDE

Slope of black-body curve 1·7% per cm.$^{-1}$ at 715 cm.$^{-1}$

Slit-width (mm.)	c (cm.$^{-1}$)	a' (cm.$^{-1}$)	A'_{max}	Triangular Slit Function		Gaussian Slit Function	
				a (cm.$^{-1}$)	A_{max} (a=9·27 cm.$^{-1}$)	a (cm.$^{-1}$)	A_{max} (a=9·2 cm.$^{-1}$)
2·00	12·5	12·0	0·78	9·3	1·00	9·7	0·985
1·80	11·3	11·6	0·80	9·45	0·99	9·7	0·975
1·50	9·5	10·95	0·835	9·25	0·97	9·55	0·97
1·00	6·5	9·85	0·89	9·15	0·96	9·15	0·96
0·60	4·1	9·5	0·925	9·25	0·955	9·2	0·96
0·40	2·9	9·35	0·935	9·2	0·945	9·2	0·955
				Av. 9·27			

It is seen that in general the calculated a value, which, if the formulae hold rigorously, should be independent of slit-width, increases somewhat for the greatest slit-widths. The increase is greater for the Gaussian than for the triangular slit function. The value of a chosen for calculating A_{max} is indicated in brackets at the top of the A_{max} column of the table. It is taken as the average of the concordant values of the set. The agreement between the values of A_{max} calculated for different slit-widths is somewhat better for the Gaussian than for the triangular slit function. It may be noted that A_{max} is not very sensitive to small variations in a.

TABLE 2

985 CM.$^{-1}$ BAND OF METHANE SULPHONYL FLUORIDE
Slope of black-body curve 0·3% per cm.$^{-1}$ at 985 cm.$^{-1}$

Slit-width (mm.)	c (cm.$^{-1}$)	a' (cm.$^{-1}$)	A'_{max}	Triangular Slit Function		Gaussian Slit Function	
				a (cm.$^{-1}$)	A_{max} (a=12·45 cm.$^{-1}$)	a (cm.$^{-1}$)	A_{max} (a=12·35 cm.$^{-1}$)
1·00	18·7	17·0	0·62	12·65	0·835	13·35	0·815
0·60	11·6	14·4	0·715	12·5	0·815	12·85	0·81
0·30	6·3	13·3	0·775	12·45	0·81	12·8	0·805
0·15	3·6	12·45	0·795	12·45	0·81	12·3	0·805
0·10	2·75	12·45	0·80	12·45	0·805	12·35	0·805

TABLE 3

1017 CM.$^{-1}$ BAND OF 1-HYDROXYANTHRAQUINONE
Slope of black-body curve 0·25% per cm.$^{-1}$ at 1017 cm.$^{-1}$

Slit-width (mm.)	c (cm.$^{-1}$)	a' (cm.$^{-1}$)	A'_{max}	Triangular Slit Function		Gaussian Slit Function	
				a (cm.$^{-1}$)	A_{max} (a=3·15 cm.$^{-1}$)	a (cm.$^{-1}$)	A_{max} (a=3·15 cm.$^{-1}$)
1·00	20·55	12·85	0·17		0·685	4·55	0·65
0·50	10·8	7·85	0·295	3·9	0·685	4·95	0·645
0·30	6·9	5·55	0·395	3·6	0·655	3·95	0·63
0·10	3·0	3·6	0·585	3·2	0·67	3·15	0·665
0·07	2·4	3·45	0·605	3·15	0·665	3·15	0·66

IV. CONCLUSIONS

It is apparent that the formulae developed in this paper, relating the true and observed breadths and intensities of infra-red absorption bands, agree well with the experimental data presented. Either a triangular or a Gaussian slit function is a fair approximation, though the calculations are less laborious with the latter. It is particularly satisfactory to note the excellent agreement between theory and experiment for the 1017 cm.$^{-1}$ band of 1-hydroxyanthraquinone, which forms a more severe test of the theory than do the other bands studied, since the spectral slit-width for most of the settings used was as great as or greater than the width of the band itself.

It seems from the three cases chosen for study that when operating the Perkin-Elmer Model 12C Spectrometer under normal conditions the band-broadening caused by finite slit-width is quite small. For the 1017 cm.$^{-1}$ band of 1-hydroxyanthraquinone for instance, with a slit-width of 0·1 mm. the band is broadened by less than 15 per cent.

V. ACKNOWLEDGMENTS

The author is indebted to Mr. A. F. Moodie for drawing his attention to the properties of the error functions and for checking some of the algebra. He also wishes to acknowledge with gratitude much discussion with Mr. C. Billington and Mr. A. Walsh.

VI. REFERENCES

AVERY, W. H., and ELLIS, C. F. (1942).—*J. Chem. Phys.* **10**: 10.

AVERY, W. H., and MORRISON, J. R. (1947).—*J. Appl. Phys.* **18**: 960.

BARNES, R. B., MCDONALD, R. S., WILLIAMS, VAN ZANDT, and KINNAIRD, R. F. (1945).—*J. Appl. Phys.* **16**: 77.

CARSLAW, H. S., and JAEGER, J. C. (1947).—"Conduction of Heat in Solids." (Oxford University Press.)

DENNISON, D. M. (1928).—*Phys. Rev.* **31**: 503.

EBERHARDT, W. H. (1950).—*J. Opt. Soc. Amer.* **40**: 172.

ELSASSER, W. M. (1938).—*Phys. Rev.* **54**: 126.

HARDY, A. C., and YOUNG, F. M. (1949).—*J. Opt. Soc. Amer.* **39**: 265.

RICHARDS, R. E., and BURTON, W. R. (1949).—*Trans. Faraday Soc.* **45**: 874.

WALSH, A., and WILLIS, J. B. (1950).—*J. Chem. Phys.* **18**: 552.

RUPTURE OF OXIDE FILMS DURING REPEATED SLIDING

By A. J. W. Moore* and W. J. McG. Tegart*

[*Manuscript received December* 20, 1950]

Summary

When copper specimens are subjected to repeated sliding, non-metallic particles are included below the surface. They are derived from the rupture of oxide films during sliding. On annealing at temperatures near the melting point of copper, the particles coalesce but prevent grain growth of the copper. The relation of these facts to the layers formed by polishing is discussed briefly.

I. Introduction

In recent years, considerable evidence has accumulated that the frictional force between sliding bodies has its origin in metallic junctions formed at the points of most intimate contact (see, for example, Bowden and Leben 1939; Ernst and Merchant 1940; and Bowden, Moore, and Tabor 1943). It has been shown (Moore 1948) that the strength of the junction is frequently as great as that of the metal itself and consequently, on sliding, rupture can occur at some distance beneath either of the two surfaces.

This paper shows that non-metallic particles are included in the surface layers as a result of the severe plastic working produced by repeated sliding.

II. Experimental

The friction experiments were carried out with an apparatus similar to that described by Bowden and Leben (1939). A hemispherical slider (approximately 6 mm. diameter) of diamond or hard steel, as required by the conditions, was caused to slide over a flat copper plate machined from O.F.H.C. copper strip. (Cu 99·98 per cent., Pb, Sn, Ni, and Fe each less than 0·005 per cent.; no copper oxide particles were detectable by microscopic examination; grain size was approximately 0·02 mm. diameter and hardness 43 D.P.H.) The load was 4000 g. and the speed of sliding 0·01 cm./sec., at which speed there is no significant rise in the temperature of the rubbing surfaces.

Clean surfaces were obtained by rubbing on a 400 grade carborundum paper under water, washing in hot running water, and drying rapidly in a stream of hot air. Such a surface is still covered by a thin oxide film and adsorbed gas or moisture layers, but is completely free from grease, and is wetted by water. Lubricated sliding was carried out using a 10 per cent. solution of oleic acid in paraffin oil.

The technique of taper sectioning was used in the microscopic examination of the surface zones. The surface was protected with a silver electrodeposit

---

* Division of Tribophysics, C.S.I.R.O., University of Melbourne.

and ground at an angle of about 5° in order to produce a surface contour with a vertical magnification of about 10 times the horizontal magnification (Moore 1948 ; Norris 1949). The specimens were etched using acid ferric chloride and ammonia-hydrogen peroxide.

III. RESULTS

There is a marked difference between the surface zones produced by the single passage of a slider under clean conditions, and the zones produced by repeated sliding in the same track. Plate 1, Figures 1 and 2, shows the areas immediately below the tracks produced after 1 " run " and 15 " runs " by a steel slider, and Plate 2, Figure 3, shows a portion of Figure 2 of Plate 1 at a higher magnification. These specimens were etched, but not annealed after sliding. The repeated sliding has produced well-marked areas exhibiting definite banding. The appearance of the bands suggests an ordered arrangement of small inclusions. When the slider had passed over the surface only twice, the same dark etching zones were observed, though they were of smaller area.

There is the possibility that the inclusions are metallic or oxide particles derived from the steel slider. However, if such is the case, a diamond slider which would not wear away should produce no banded zone. Plate 2, Figure 4, shows the copper after sliding 15 times with diamond. The dark etching areas are still present. Therefore they are derived from the copper. The evidence suggests that the inclusions are cuprous oxide which has been formed on the surface after each sliding and then moved below the surface during the intense deformation which accompanies the next passage of the slider.

When the specimens are annealed subsequent to sliding, coalescence of the inclusions occurs. Plate 3, Figures 5 and 6, shows the tracks produced after sliding 15 times with steel, and then annealing at 700 and 900 °C. respectively in a vacuum for 2 hours ; Plate 3, Figure 7, shows a portion of Figure 6 at a higher magnification. It will be seen from these photomicrographs that although the copper was subjected to prolonged annealing near its melting point and extensive grain growth has occurred, no such grain growth has taken place in the immediate neighbourhood of the coalesced particles. Here the grains have not grown beyond about 0·005 mm. diameter. With lubricated sliding, no evidence of the inclusions has been found. Moreover, when the tracks made by 15 repeated lubricated " runs " of steel are annealed at 500 °C., the microstructure shows that grain growth has not been inhibited, since large grains occur right up to the surface of the track (Plate 4, Fig. 8).

Some experiments were carried out in which repeated sliding took place in the presence of dilute hydrochloric acid with the object of preventing the (re-)formation of the protective oxide film. The friction was very high, the average coefficient of friction for 15 " runs " was 1·22, varying between 0·86 and 2·04. A section (Plate 4, Fig. 9) shows that although intense deformation has occurred at least 15μ below the surface, the proportion of inclusions is less than that which results from sliding without hydrochloric acid present. The reduction in the proportion of inclusions is less than with the experiments with lubricant; in the latter case their formation is inhibited by the lubricant film both preventing the formation of the oxide film and reducing the extent of surface deformation.

IV. Discussion

Difficulties encountered when the general concept, that friction is due to the shearing of metallic junctions, is applied to specific cases have recently been explained in terms of the behaviour of oxide films which are known (Miley and Evans 1937) to form rapidly at room temperature.

Whitehead (1950) has shown that for some metals the coefficient of friction decreases with load over a load range 10^{-1} to 1 g., becoming constant again at lower loads. Here the slight deformation occurring at junctions under low loads is not sufficient to break the relatively brittle oxide film, and thus sliding occurs between oxides. Moore and Tegart (unpublished data) have shown that for copper-beryllium alloys, friction decreases approximately linearly with hardness, a result which is not expected from the general theory. This can be explained in terms of critical strain of the substrate before rupture of the oxide occurs.

Again the results of Bowden and Young (1949) on friction *in vacuo* and McFarlane and Tabor (1950) on the relation between friction and adhesion are explained similarly.

The experiments described in this paper illustrate the role played by the oxide film when sliding is continued. After rupture of the oxide at points of critical plastic flow, adhesion occurs and the resulting deformation in the surface layers is sufficiently severe for oxide particles to be included. On further sliding, successive layers of oxide are broken up and entrapped.

Since the non-metallic particles inhibit grain growth during sliding, the surface layers are able to remain stable at a degree of deformation which is higher than that generally produced by other means. The rubbing surfaces encountered in many practical applications have been altered in this way, a fact which may influence their lubricating properties.

The remarkable inhibiting effect of the particles on grain growth is probably also important in the production of a polished layer on metal surfaces. If the distance between the particles is very small, as may be expected during polishing where no coalescence could occur because the heating would be localized and momentary, the grain size of the metal must remain small. Conditions of polishing are, therefore, such that finely crystalline or even amorphous metal could be formed and remain stable.

V. References

Bowden, F. P., and Leben, L. (1939).—*Proc. Roy. Soc.* A **169**: 371.

Bowden, F. P., Moore, A. J. W., and Tabor, D. (1943).—*J. Appl. Phys.* **14**: 80.

Bowden, F. P., and Young, J. E. (1949).—*Nature* **164**: 1089.

Ernst, H., and Merchant, M. E. (1940).—Conference on friction and surface finish. Mass. Inst. Technol. p. 76.

McFarlane, J. S., and Tabor, D. (1950).—*Proc. Roy. Soc.* A **202**: 244.

Miley, H. A., and Evans, U. R. (1937).—*Nature* **139**: 283.

Moore, A. J. W. (1948).—*Proc. Roy. Soc.* A **195** : 231.
Moore, A. J. W. (1948).—*Metallurgia* **38** : 71.
Norris, D. J. (1949).—Commonw. Dep. Supply and Development Div. Aeronaut. S. & M. Note 176.
Whitehead, J. R. (1950).—*Proc. Roy. Soc.* A **201** : 109.

Explanation of Plates 1–4

Plate 1

Fig. 1.—Section below track after one passage of a steel slider. Not annealed after sliding. ×125.

Fig. 2.—Section below track after a steel slider has passed over the surface 15 times. Not annealed after sliding. ×125.

Plate 2

Fig. 3.—Portion of Plate 1, Figure 2, at a higher magnification. ×550.

Fig. 4.—Section below track after a diamond slider has passed over the surface 15 times. ×100.

Plate 3

Fig. 5.—Section below track after a steel slider has passed over the surface 15 times. Annealed at 700 °C. after sliding. ×110.

Fig. 6.—Same as Figure 5, but annealed at 900 °C. ×110.

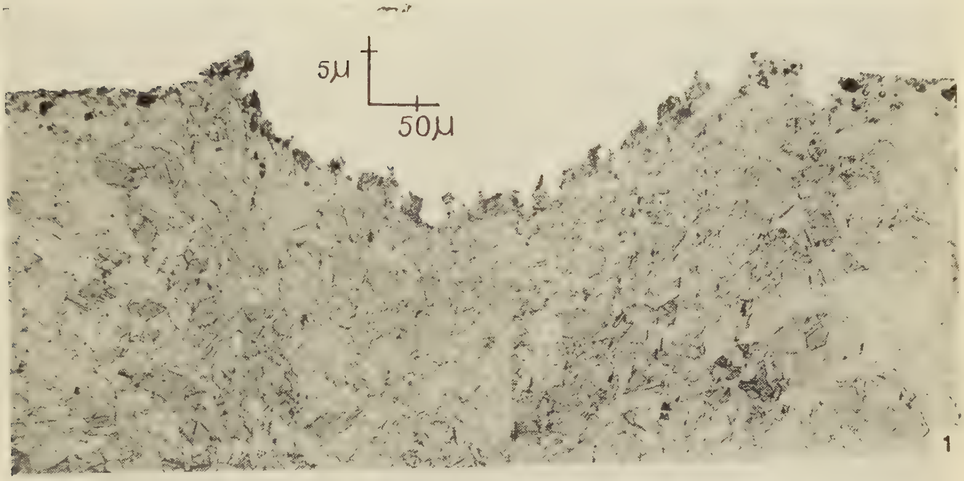
Fig. 7.—Portion of Figure 6 at a higher magnification. ×600.

Plate 4

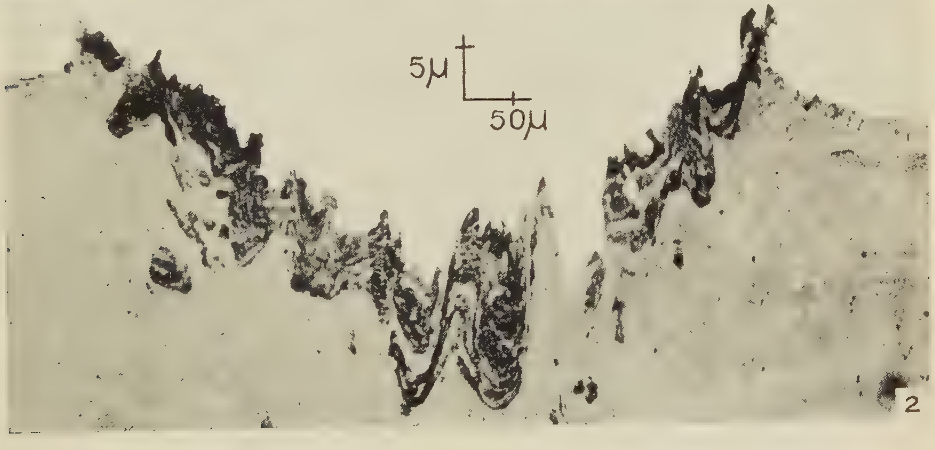
Fig. 8.—Section below track after a steel slider has passed over the surface 15 times under lubricated conditions. Annealed at 500 °C. after sliding. ×125.

Fig. 9.—Section below track after sliding 15 times with a steel slider under dilute hydrochloric acid. ×125.

RUPTURE OF OXIDE FILMS DURING REPEATED SLIDING

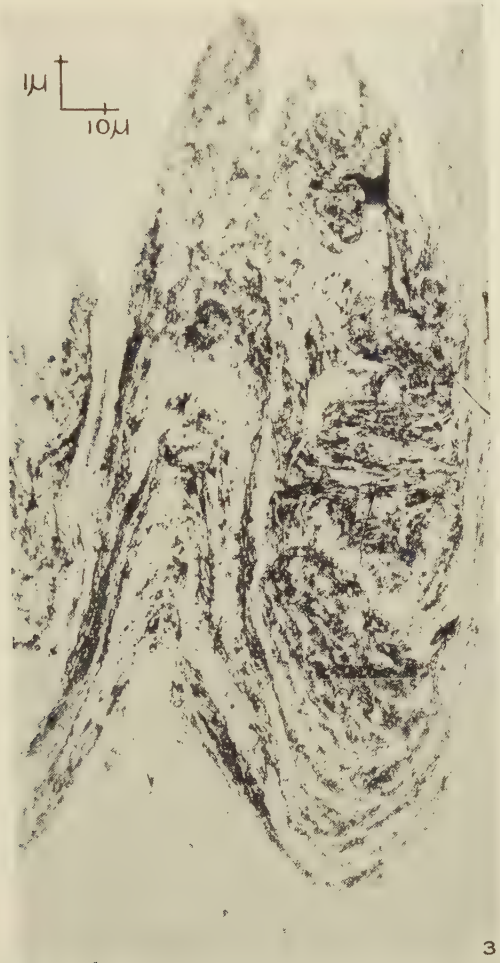


1

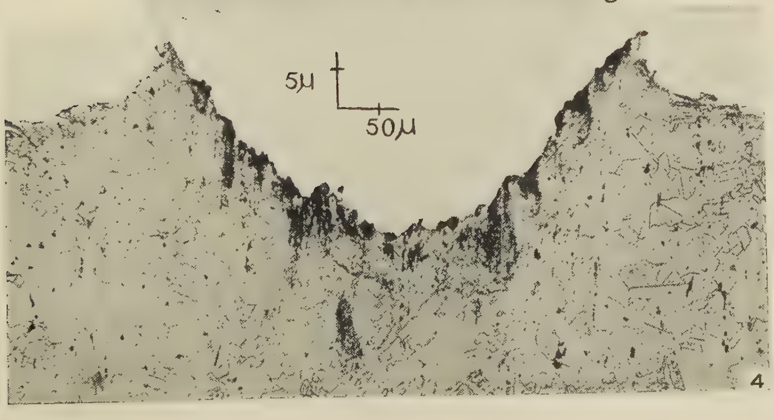


2

RUPTURE OF OXIDE FILMS DURING REPEATED SLIDING

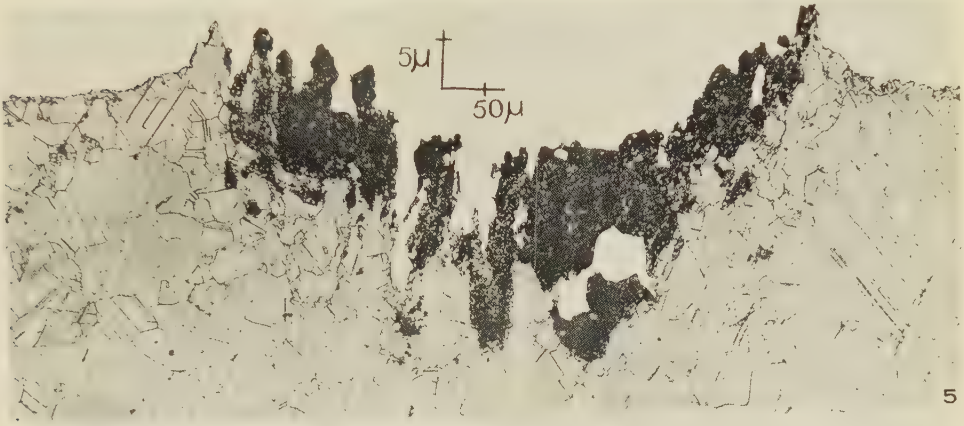


3

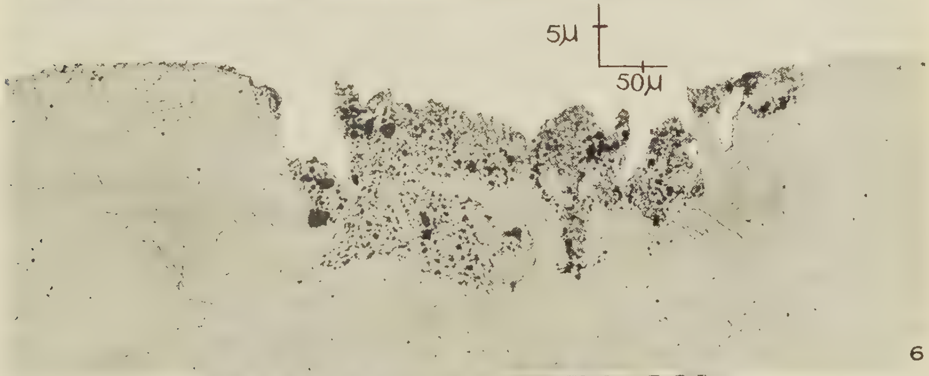


4

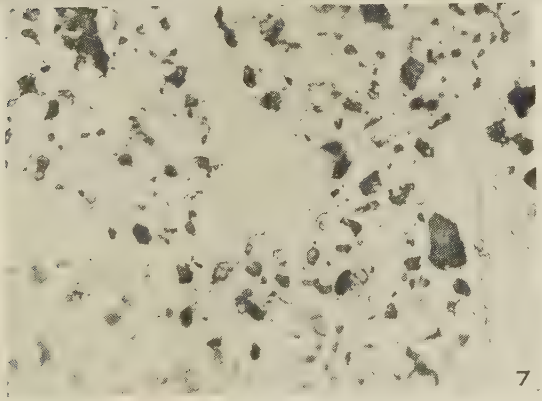
RUPTURE OF OXIDE FILMS DURING REPEATED SLIDING



5

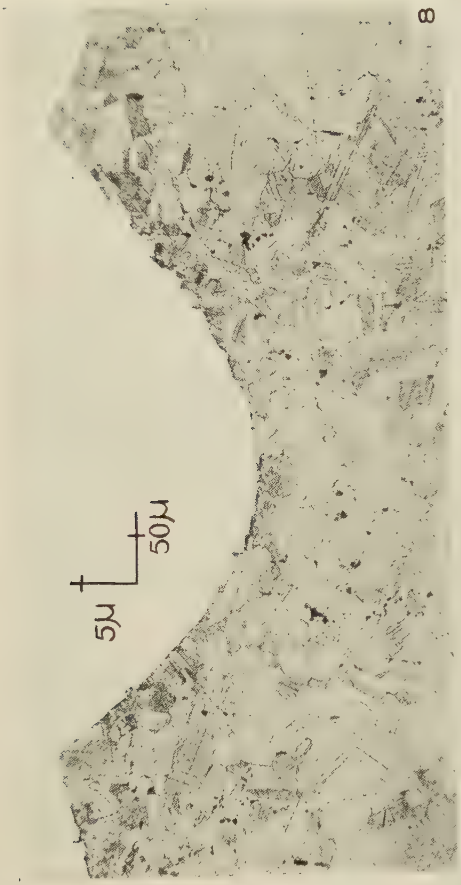


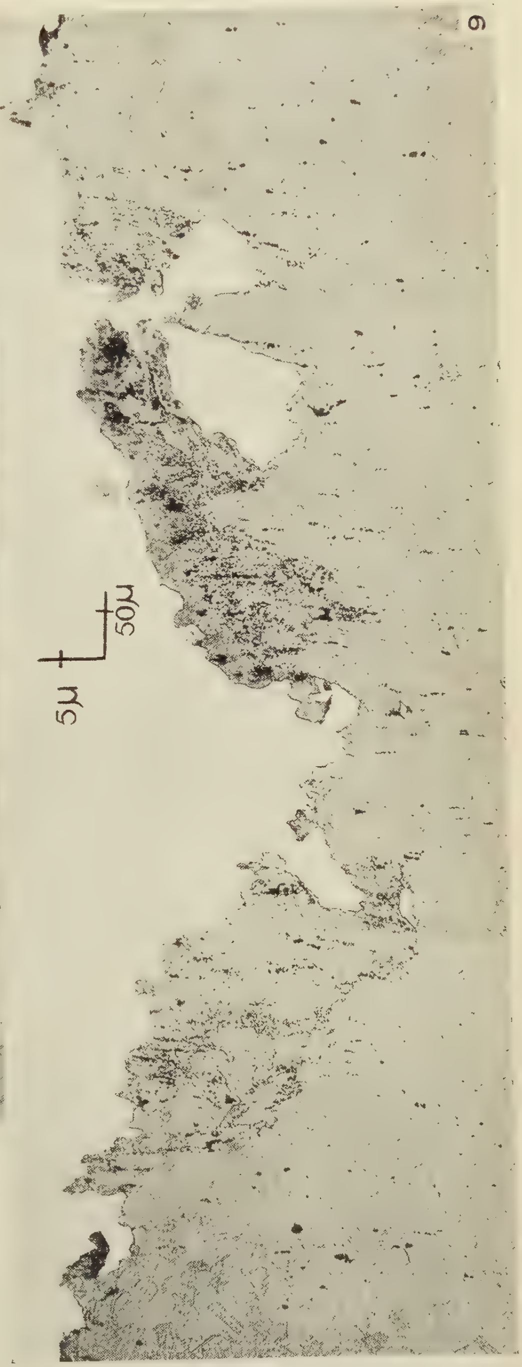
6



7

RUPTURE OF OXIDE FILMS DURING REPEATED SLIDING


8


9

STUDIES OF THE WALLACH TRANSFORMATION

II. STERIC EFFECTS IN SOME SYMMETRICALLY DISUBSTITUTED AZOXYBENZENES

By P. H. Gore* and G. K. Hughes*

[*Manuscript received October* 11, 1950]

Summary

Study of the behaviour of symmetrical azoxy compounds in the Wallach Transformation† shows the important influence of size of a substituent group *ortho* to the point of entry of the hydroxyl group.

The various processes known to occur in a solution of azoxybenzene in sulphuric acid are discussed.

I. Introduction and Results

3,3′-Dimethylazoxybenzene was found by Parsons and Bailar (1936) to rearrange to 3,3′-dimethyl-4-hydroxyazobenzene, but no mention was made of by-products. This substance and the 2,2′- and 4,4′-isomers have now been subjected to the W.T. under identical conditions (Expts. IV–VI)‡ and results obtained are shown in Table 1.

Table 1

Azoxybenzene	Hydroxyazo Compound (%)	Azo Compound (%)	Polymer (%)	Sulphonic Acids etc. (%)	Total Phenols (%)
2,2′-Dimethyl	31	43	5	21	[52]
3,3′-Dimethyl	50–55	33	12–14	2–3	[52–58]
4,4′-Dimethyl	47	22	23	7–8	[54–55]

The W.T. of 3,3′-dimethylazoxybenzene shows an unusual feature. The fraction which normally contains only the hydroxyazo compound(s) in this instance has an appreciable amount of resinous semi-solid impurities. These were not *o*-hydroxyazo compounds, as steam distillation slowly produced nearly pure 3,3′-dimethyl-4-hydroxyazobenzene, m.p. 110–111 °C. (Jacobsen *et al.* (1895) give m.p. 115 °C.; Lukashevish and Khurdyumova (1949) give m.p. 108–110 °C.; 3,3′-dimethyl-2-hydroxyazobenzene (?), m.p. 61·5–62·5 °C., Cumming and Ferrier (1925); 5,3′-dimethyl-2-hydroxyazobenzene, m.p. 95 °C.,

---

* School of Chemistry, University of Sydney.

† Hereafter abbreviated to W.T.

‡ Results for 4,4′-dimethylazoxybenzene are taken from Gore and Hughes (1950).

Jacobsen and Piepenbrinck (1894)). The impurities, owing to their solubility in ether, are regarded as low molecular weight polymers with small sulphonic acid content.

Since quantitative separation of 3,3′-dimethyl-4-hydroxyazobenzene from the resinous polymer proved impracticable (chromatographic adsorption on alumina, fractional crystallization, or precipitation having failed), attempts were made to determine the yield by indirect means. A W.T. was conducted on 3,3′-dimethylazoxybenzene under the same conditions as before, except for 3½ hours instead of ½ hour, during which time most of the semi-solid material was converted to a " normal " polymer. To extrapolate to time ½ hour, the effect of sulphuric acid on 3,3′-dimethyl-4-hydroxyazobenzene (other products of the W.T. are reasonably stable in sulphuric acid, cf. Gore and Hughes 1950) for 3 hours' duration, was measured and found to give an unusually high yield (55%) of a vermilion crystalline " polymer " (Expt. VII) (resembling those formed from 4-hydroxyazobenzene and 2,2′-dimethyl-4-hydroxyazobenzene under similar conditions), which is being further investigated. In consequence this method was discarded.

The behaviour of the three symmetrical dichloroazoxybenzenes has also been investigated under standard conditions (Table 2).

TABLE 2

Azoxybenzene	Hydroxyazo Compound (%)	Azo Compound (%)	Polymer (%)	Sulphonic Acids etc. (%)	Total Phenols (%)
2,2′-Dichloro	54	36	1·1	8–9	[62–63]
3,3′-Dichloro	79	17	Trace	3–4	[82–83]
4,4′-Dichloro	11	74	1·3	13–14	[24–25]

With 3,3′-dichloroazoxybenzene, Schültz (1884) reported the formation of a faint trace of 3,3′-dichloroazobenzene and a near quantitative yield of 3,3′-dichlorohydroxyazobenzene, m.p. 114–115 °C., but did not propose a structure. Richter (1900) (cf. Knipscheer 1903) claimed this to be the 4-hydroxy isomer, but gave no proof. This has now been verified by comparison with an authentic specimen of 3,3′-dichloro-4-hydroxyazobenzene, m.p. 125–126 °C.

Heumann (1872) reported the formation of 4,4′-dichloroazobenzene from 4,4′-dichloroazoxybenzene, as well as considerable quantities of colouring matters and non-crystallizable bodies. Schultz (1884) later found a trace of a hydroxyazo compound, which was not further investigated. This (11%, m.p. 165–167 °C.) has now been shown to be 4,4′-dichloro-2-hydroxyazobenzene by reduction to the known 2-amino-5-chlorophenol. Cumming and Ferrier (1925) reported the formation of a phenol, m.p. 171–172 °C., from the action of light on 4,4′-dichloroazoxybenzene, but did not suggest a structure; it is most probably the same as the above.

In the W.T. of 4,4′-dichloroazoxybenzene the yield of polymer (determined in the usual way) was only 1·3 per cent., with a yield of soluble sulphonic acids of 13–14 per cent. The clear mother liquor, on boiling, deposited additional quantities of polymer (6·8%) during several hours (cf. Expt. VI). It thus appears that polymer can be slowly produced from the sulphonic acids present in the mother liquor,* as well as from the hydroxyazobenzenes and the main source, the azoxybenzenes (cf. Gore and Hughes 1950).

Comparative experiments were also conducted with the 3,3′-dihalogenoazoxybenzenes and the parent substance, in order to determine the effect of increasing substituent size on the yield of hydroxyazo compound (Expts. II, VIII–XI). Results are shown in Table 3.

Table 3

Azoxybenzene	Hydroxyazo Compound (%)	Azo Compound (%)	Polymer (%)	Sulphonic Acids etc. (%)	Total Phenols (%)
3,3′-Difluoro-	65	14	Trace	20–21	[85–86]
3,3′-Dichloro- ..	79	17	Trace	3–4	[82–83]
3,3′-Dibromo-	47	52	Trace	1	[48]
3,3′-Diiodo-	17	71	3	9	[26]

3,3′-Difluoroazoxybenzene was prepared by reduction of 3-fluoronitrobenzene by magnesium in methanol. This gave, somewhat unexpectedly, a molecular compound of 3,3′-difluoroazobenzene with 3,3′-difluoroazoxybenzene, which on oxidation gave the required azoxy compound. The structure of the hydroxyazo compound formed in the W.T. of 3,3′-difluoroazoxybenzene is believed to be 3,3′-difluoro-4-hydroxyazobenzene by analogy with the phenols from 3,3′-dichloroazoxybenzene (*vide supra*) and from 3,3′-dibromoazoxybenzene, which was shown to be 3,3′-dibromo-4-hydroxyazobenzene by direct comparison with an authentic specimen.

The phenolic product from the W.T. of 3,3′-diiodoazoxybenzene was a resinous mixture, from which two crystalline benzoyl derivatives were, however, obtained. The lower melting substance, $C_{19}H_{12}O_2N_2I_2$, was shown (by direct comparison with an authentic specimen) to be the benzoyl derivative of 5,3′-diiodo-2-hydroxyazobenzene ; the higher melting substance was apparently a mixture containing a triiodobenzoyloxyazobenzene, which was formed by an acid catalyzed disproportionation of the primary rearrangement product (Neumann 1887 ; Nicolet 1921) together with a conjugate product, presumably 3-iodo-2′-hydroxyazobenzene—this will be discussed more fully elsewhere.

* No polymer formation could be detected in the action of 98 per cent. sulphuric acid on the potassium salt of 4-hydroxyazobenzene-4′-sulphonic acid (Expt. XI), which according to Lachman (1902) is the main sulphonic acid from azoxybenzene.

II. Discussion

(i) The various reactions that are now known to occur in a reaction mixture of azoxybenzene with sulphuric acid are summarized in Figure 1.

Reactions 1–4 are fast, 5 and 6 are slow, 7 is very slow ; 5 may be a combination of 4 and 6. Usually 1 is faster than 3, and 2 than 4.

The action of sulphuric acid on azoxybenzene accounts for most of the azobenzene (>95%) (Gore and Hughes (1950) ; cf., however, Lachman (1902), who reported yields of up to 16 per cent. from 4-hydroxyazobenzene by the action of monohydrate), all of the hydroxyazobenzenes, and most of the polymer (>92%). The sulphonic acids are probably formed exclusively from the hydroxyazo compounds, or a common precursor.

(ii) From Tables 1 and 2 it can be seen that a maximum yield of hydroxyazo compound in each series is obtained with the 3,3′-disubstituted azoxybenzene. It was, however, shown in Part I of this series (Gore and Hughes 1950) that the yield of hydroxyazobenzene goes through a maximum, and its ultimate yield depends not only on the primary yield but on its rate of removal (i.e. by sulphonation). Unfortunately the rates of sulphonation of the three dimethylhydroxyazobenzenes could not be determined (*vide supra*, cf. Expt. VII), since these reactions are anomalous. Since, however, it is believed that the sulphonic acids are formed nearly entirely from the hydroxyazo compounds, their sum is a more accurate gauge. Table 1 shows that there is no appreciable difference between these three isomers within experimental error. The case of the symmetrical dichloroazoxybenzenes is more clear-cut, a maximum occurring with the 3,3′-isomer (Table 2).

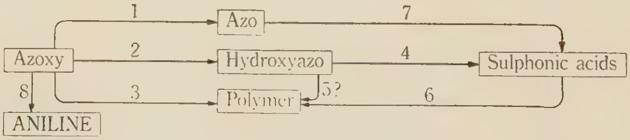


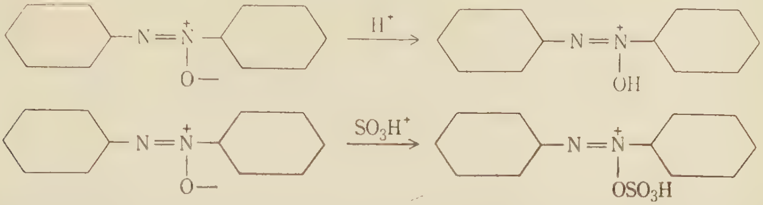
Fig. 1

From Table 4 it can be seen that an azoxybenzene molecule containing electron-attracting groups (nitro or halogeno) gives an improved yield of total phenols compared to the control experiment of azoxybenzene (Expt. XV), whereas electron-donating groups (cf. also methoxy, Gore and Hughes 1950) cause a lowered yield.

(iii) The experiments with halogen substituted azoxybenzenes show clearly that steric effects play a very important part in determining the yields of hydroxyazo compound. Subsequent arguments* of the steric effect of substituent groups are based on the reasonable assumption that in strongly acid medium, such as concentrated sulphuric acid, an azoxy compound will react with protons, or possibly another cation such as SO_3H^+.

---

* It is not intended here to give a detailed discussion of the mechanism.



The addition of a proton would not be affected appreciably by the proximity of substituent groups, whilst that of SO_3H^+ would.

Table 4

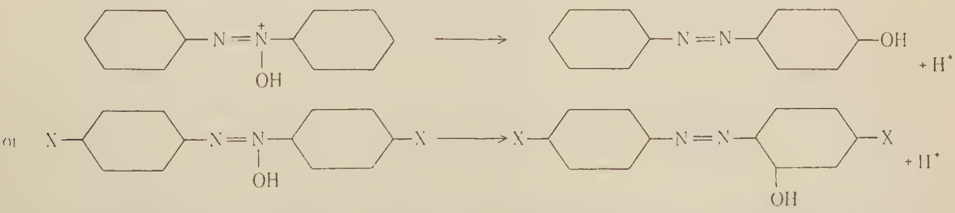
REARRANGEMENT OF 3,3′-X_2-AZOXYBENZENES

Conditions: (*a*) 98% H_2SO_4, 90 °C., 1 hr.; (*b*) 85% H_2SO_4, 90 °C., $\frac{1}{2}$ hr.; (*c*) 98% H_2SO_4, 140 °C., 1·8 hr.

X	van der Waals Radius of X (A)	Hydroxyazo Compound (%)	Total Phenols (%)	Conditions	Reference
F ..	1·35	65	[85]	(*a*)	Expt. VIII
Cl ..	1·80	79	[82–83]	(*a*)	Expt. II
Br ..	1·95	47	[48]	(*a*)	Expt. XI
I ..	2·15	17	[26]	(*a*)	Expt. X
H ..	1·2	3	[45–46]	(*a*)	Gore & Hughes (1950)
CH_3 ..	2·0	50–55	[52–58]	(*b*)	Expt. VI
H ..	1·2	62	[70]	(*b*)	Gore & Hughes (1950)
NO_2 ..	—	56	[79]*	(*c*)	

* This value is too high, the "sulphonic acid" fraction containing in this case considerable amounts of amine.

The rearrangement of the charged molecule must involve at least two further steps: (*a*) removal of the –OH or –OSO_3H group (radical or ion) (the latter can also be formed by direct sulphation of the –OH group), and (*b*) introduction of an –OH or –OSO_3H group in the *p*- or *o*-positions.



It is at present unknown which of these processes is actually the first to occur. If (*b*) is the first, the W.T. cannot be intramolecular, and if (*a*) is the first, the W.T. may be intra- or intermolecular.

Bulky substituent groups in the *o*-positions of the azoxybenzene molecule would sterically hinder formation of the $-OSO_3H$ group, and removal of the $-OH$ or $-OSO_3H$ group if the reaction is bimolecular, i.e. step (*a*), during the rearrangement. Similarly, in the case of the 4,4′-disubstituted azoxybenzene, the benzeneazo group could sterically hinder approach to the *o*-positions, i.e. step (*b*). (If, however, the rearrangement is intramolecular, there is no apparent reason why an *o*-hydroxyazo compound should not be formed in good yield in the case of 4,4′-disubstituted azoxybenzenes. The yield with 4,4′-dimethylazoxybenzene is 47 per cent., but with 4,4′-dichloroazoxybenzene only 11 per cent.) Furthermore, step (*b*) should also be hindered in a 3,3′-disubstituted azoxybenzene. On this basis one would expect a higher yield of hydroxyazo compound with an azoxybenzene with small substituent groups in 3,3′-positions.

Table 4 clearly shows that this is the case with the 3,3′-dihalogenoazoxybenzenes, the yield of phenols decreasing rapidly with increased size of the substituent group. With a very large group, such as iodine, the hydroxyl group does not enter the 4-position at all, preferring the sterically more accessible 6-position. This is the first case of the W.T. giving an *o*-hydroxy compound with an azoxy compound having an unsubstituted *p*-position.

From Expts. XII and XIII it is seen that 3,3′-difluoro-4-hydroxyazobenzene sulphonates at a considerably greater rate than 3,3′-dichloro-4-hydroxyazobenzene ; this accounts for the lower primary yield of the former substance, and was to be expected from the greater inductive effect of fluorine (Schwarzenbach and Michaelis 1938).

It also follows (Expts. II, IX, XIII, and XIV) that rearrangement of 3,3′-dichloroazoxybenzene is more rapid than of 3,3′-dibromoazoxybenzene.

III. EXPERIMENTAL

(*a*) *2,2′-Dichloroazoxybenzene*

TABLE 5

Expt.	Conditions	Hydroxyazo Compound	Azo Compound	Polymer	Sulphonic Acids etc.
I	4·8 g., 50 ml. of 98%, 90 °C., 1 hr.	54%	36%	1%	9%

The hydroxyazo compound crystallized from aqueous ethanol in lustrous brown needles, m.p. 182–182·5 °C.

Found : N, 10·4%.

Calculated for $C_{12}H_8ON_2Cl_2$: N, 10·5%.

It was reduced with stannous chloride and hydrochloric acid by heating on the water-bath for 15 minutes. The pale yellow solution was filtered, made alkaline, and *o*-chloraniline removed by ether extraction. After careful neutralization, ether extraction removed the aminophenol, m.p. 158·5–159·5 °C. (crude), 160 °C. (pure) ; the m.p. of 3-chloro-4-aminophenol is 160 °C., German Patent 143449 (1903). The hydroxy compound was therefore 2,2′-dichloro-4-hydroxyazobenzene.

The azo compound had m.p. 129–135 °C. (crude), 135–136 °C. after recrystallization from ethanol, and was therefore 2,2′-dichloroazobenzene (m.p. 136 °C., Brand 1903).

(*b*) *3,3′-Dichloroazoxybenzene*

TABLE 6

Expt.	Conditions	Hydroxyazo Compound	Azo Compound	Polymer	Sulphonic Acids etc.
II	4·8 g., 50 ml. of 98%, 90 °C., 1 hr.	79%	17%	Trace	3–4%

The phenol had m.p. 122–124 °C. (crude), 125–126 °C. (from aqueous ethanol) and was identical in all respects to 3,3′-dichloro-4-hydroxyazobenzene synthesized below (Schultz (1884) gives m.p. 114–115 °C.).

Found: N, 10·3%.

Calculated for $C_{12}H_8ON_2Cl_2$: N, 10·3%.

The azo compound crystallized from ethanol in lustrous orange needles, m.p. 104 °C. (101 °C., Laubenheimer 1875; Vorländer and Meyer 1902).

3,3′-Dichloro-4-hydroxyazobenzene and 2-Chloro-4,6-bis(3′-chlorobenzeneazo)-phenol.—*m*-Chloraniline (5 g.) was diazotized in the usual way, and coupled with *o*-chlorophenol (5 g.) in alkaline solution. After 2 hours, some brown bisazo compound* was removed by filtration, and the mixture of hydroxyazo compounds precipitated by acid. This was filtered, washed, and taken up in boiling 50% acetic acid, which on cooling deposited small brown needles, m.p. 178–180 °C. (Yield 0·94 g., 6% calculated as bisazo compound.)

2-Chloro-4,6-bis(3′-chlorobenzeneazo)-phenol was recrystallized from glacial acetic acid, m.p. 183–184 °C.

Found: N, 13·8%.

Calculated for $C_{18}H_{11}ON_4Cl_3$: N, 13·8%.

It was insoluble in cold, slightly soluble in hot 10% sodium hydroxide solution, and soluble in hot 1% sodium hydroxide solution giving a red solution.

On addition of water to the hot mother liquor above, and cooling, brown needles of 3,3′-dichloro-4-hydroxyazobenzene, m.p. 125–126 °C., separated (7·8 g., 74%).

(*c*) *4,4′-Dichloroazoxybenzene*

The hydroxyazo compound, m.p. 158–163 °C. (crude), crystallized from aqueous acetic acid in yellow microscopic needles, m.p. 165–167 °C.

Found: Cl, 26·0%.

Calculated for $C_{12}H_8ON_2Cl_2$: Cl, 26·5%.

Reduction of this phenol, by stannous chloride-hydrochloric acid mixture as above, yielded brown crystals of an aminophenol, m.p. 153–154 °C. (2-Amino-5-chlorophenol has m.p. 154 °C., Hodgson and Kershaw 1928), which must have been derived from 4,4′-dichloro-2-hydroxyazobenzene.

---

* The formation of a bisazo along with the monoazo compound using less than 1 mole of diazotized amine in the coupling reaction (cf. Experimental) seems to be a general behaviour (cf. Gore and Hughes 1949).

The azo compound, m.p. 175–179 °C. (crude), crystallized from ethanol in long silky orange needles, m.p. 184 °C. (m.p. of 4,4′-dichloroazobenzene ranges from 183 °C., Heumann 1872, to 187·5–188·5 °C., Bamberger and Baudisch 1909).

TABLE 7

Expt.	Conditions	Hydroxyazo Compound (%)	Azo Compound (%)	Polymer (%)	Sulphonic Acids etc. (%)
III	7·0 g., 72 ml. of 98%, 90 °C., 1 hr.	11	74	(i) 1·3 (ii) 1·3 (iii) 3·6 (iv) 0·6	8–9

Polymer (I) was the "normal" W.T. polymer, whilst polymers (II), (III), and (IV) were formed by boiling the mother liquor containing sulphonic acids (yield 13–14%) for consecutive 4 hour (approx.) periods, at concentrations of sulphuric acid *c.* 6N, 12–15N, and 12–15N, respectively. During the boiling a pronounced phenolic odour was noticeable. The nature of polymers (II)–(IV) (yield taken w/w) was not otherwise investigated.

(*d*) *2,2′-Dimethylazoxybenzene*

TABLE 8

Expt.	Conditions	Hydroxyazo Compound	Azo Compound	Polymer	Sulphonic Acids etc.
IV	10 g., 100 ml. of 85%, 90 °C., ½ hr.	31%	43%	5%	21%

(*e*) *3,3′-Dimethylazoxybenzene*

Preparation.—A dilute solution of 3,3′-dimethylazobenzene in glacial acid was treated with excess 30% hydrogen peroxide at 90 °C. for 3 hours (cf. Angeli 1910). The solution was evaporated to small bulk, diluted with water, and cooled. Yield 98%, m.p. (crude) 35–36 °C. (m.p. 37 °C., Bamberger and Renauld 1897).

Polymer V was red-brown. Found: N, 7·1%.

The reaction mixture (Expt. VI) was poured onto crushed ice (400 g.), and sufficient alcohol (*c.* 200 ml.) added to take up the resinous precipitate that had formed. The dark mixture was thoroughly ether extracted after dilution to 1 l. The aqueous liquor was filtered to remove "normal" polymer (yield 2·1%; found: N, 7·2%). The red-brown ether extract was separated in the usual way into alkali insoluble and soluble fractions. The former afforded 3,3′-dimethylazobenzene (yield 3·1 g., 33%), the latter a near black resinous viscous syrup (yield *c.* 65% w/w), which could not be satisfactorily crystallized. Steam distillation slowly removed nearly pure 3,3′-dimethyl-4-hydroxyazobenzene (crystals, 0·25 g., 8·5%, separating: m.p. 110–111 °C.). On passing an ethanolic solution of the mixture through a column of alumina, some normal polymer was strongly adsorbed and recovered by elution with 4% sodium hydroxide solution (yield 0·65 g., 9·7%), the filtrate leaving on evaporation a resinous mixture similar to the above. Microscopic examination of the solid, m.p. 94–104 °C., precipitated from an alkaline solution, showed it to consist of slender needles and dark brown globules. Both fractional precipitation and crystallization proved unsuccessful in separating the pure hydroxyazo compound.

TABLE 9

Expt.	Conditions	Hydroxyazo Compound (%)	Azo Compound (%)	Polymer (%)	Sulphonic Acids etc. (%)
V	10 g., 100 ml. of 85%, 90 °C., 3½ hr.	29	32	24	15
VI	10 g., 100 ml. of 85%, 90 °C., ½ hr.	*c.* 52	33	*c.* 13	2–3

The first mother liquor on standing for 36 days deposited another crop of polymer (0·05 g.).

Action of Sulphuric Acid on 3,3′-Dimethyl-4-hydroxyazobenzene.—On addition of the hydroxyazo compound to 85% sulphuric acid, a bulky orange precipitate (probably an addition compound) was rapidly formed, which slowly dissolved on heating the mixture.

TABLE 10

Expt.	Conditions	Hydroxyazo Compound	Polymer	Sulphonic Acids
VII	6·0 g., 100 ml. of 85%, 90 °C., 3 hr.	33%	55%	12%

The "polymer" (found: N, 9·2%) formed in this reaction was a vermilion crystalline powder; on heating to 360 °C. it darkened to a near black, losing its crystallinity.

(*f*) *3,3′-Difluoroazoxybenzene*

Prepared from *m*-fluoronitrobenzene (Shiemann and Pillarsky 1929) by the general method of Zechmeister and Rom (1929): To a solution of *m*-fluoronitrobenzene (2·76 g.) in methanol (140 ml.), clean magnesium turnings (3·2 g.) were added, and the mixture brought to boiling on the water-bath. The source of heat was removed when a vigorous reaction had set in; after 30 minutes, the mixture was again heated until solution of the magnesium was complete. The magnesium complex was decomposed by addition of 10E hydrochloric acid (60 ml.). The solution was diluted with water, cooled, and the product collected. The difluoroazodifluoroazoxy addition compound was obtained as orange needles from aqueous methanol, yield 1·51 g., 68%, m.p. 46 °C.

Found: C, 63·5, 63·2; H, 3·6, 3·6; N, 12·4%.

Calculated for $C_{24}H_{16}ON_4F_4$: C, 63·7; H, 3·6; N, 12·4%.

Oxidation of this in glacial acetic acid at 90 °C. with excess 30% hydrogen peroxide, dilution with water, and cooling produced 3,3′-difluoroazoxybenzene in quantitative yield. Lustrous lemon-yellow flakes from aqueous methanol, m.p. 51·5 °C.

Found: C, 61·3; H, 3·5%.

Calculated for $C_{12}H_8ON_2F_2$: C, 61·5; H, 3·4%.

The hydroxyazo compound crystallized in yellow-brown flakes from acetic acid-ethanol mixture, m.p. 135–136·5 °C.

Found : N, 12·1%.

Calculated for $C_{12}H_8ON_2F_2$: N, 12·0%.

TABLE 11

Expt.	Conditions	Hydroxyazo Compound	Azo Compound	Polymer	Sulphonic Acids etc.
VIII	0·614 g., 7·3 ml. of 98%, 90 °C., 1 hr.	65%	14%	Trace	20–21%

The benzoyl derivative of this (Schotten-Baumann) crystallized in light orange tufts from aqueous ethanol, or aqueous acetic acid, m.p. 107·5–108 °C.

Found : N, 8·6%.

Calculated for $C_{19}H_{12}O_2N_2F_2$: N, 8·3%.

The azo compound formed orange-red crystalline flakes from aqueous ethanol or aqueous acetic acid, m.p. 71 °C.

Found : N, 12·8%.

Calculated for $C_{12}H_8N_2F_2$: N, 12·8%.

(*g*) *3,3′-Dibromoazoxybenzene*

TABLE 12

Expt.	Conditions	Hydroxyazo Compound	Azo Compound	Polymer	Sulphonic Acids etc.
IX	3·2 g., 25 ml. of 98%, 90 °C., 1 hr.	47%	52%	Trace	1%

The hydroxyazo compound formed yellow-brown crystals from aqueous acetic acid, m.p. 124 °C., and did not depress the m.p. of 3,3′-dibromo-4-hydroxyazobenzene, synthesized as below.

Found : N, 8·2%.

Calculated for $C_{12}H_8ON_2Br_2$: N, 7·9%.

The benzoyl derivative (Schotten-Baumann) formed small orange crystals from acetic acid-ethanol mixture, m.p. 149–150 °C.

Found : N, 6·4%.

Calculated for $C_{19}H_{12}O_2N_2Br_2$: N, 6·2%.

The azo compound had m.p. 124·5–125·5 °C. (m.p. 125·5 °C. for 3,3′-dibromoazobenzene, Gabriel 1876).

3,3′-Dibromo-4-hydroxyazobenzene.—Preparation and separation as for corresponding chloro compounds.

Yield of " bisazo compound " : 14%. Crystallized from pyridine or 80% acetic acid in brown microscopic needles, m.p. 197–198 °C. Found : N, 12·1%.

The nature of this product was not otherwise investigated. It was insoluble in hot or cold 10%, and slightly soluble in 3% sodium hydroxide solution; insoluble in common organic solvents except dioxan and nitrobenzene, slightly soluble in butanol and acetic acid.

Yield of 3,3′-dibromo-4-hydroxyazobenzene was 60%, pure after three crystallizations from aqueous acetic acid.

(*h*) *3,3′-Diiodoazoxybenzene*

Improved preparation by method of Zechmeister and Rom (1929), cf. Subsection (*f*). Crude yield: 91%. Recrystallized from glacial acetic acid, m.p. 120·5–121·5 °C. (m.p. 118–119 °C., Gabriel 1876).

TABLE 13

Expt.	Conditions	Hydroxyazo Compound	Azo Compound	Polymer	Sulphonic Acids
X	4·05 g., 25 ml. of 98%, 90 °C., 1 hr.	17%	71%	3%	9%

During the progress of the reaction, traces of iodine vapour could be detected.

The azo compound had m.p. 149–150 °C. (m.p. of 3,3′-diiodoazobenzene is 150–151 °C., Meldola and Andrews 1896).

The hydroxyazo compound fraction (0·69 g.) was a dark resinous product, m.p. from 40 °C., which could not be purified by recrystallization. The mixture was benzoylated (Schotten-Baumann) giving a crude yield of 0·55 g., corresponding to 0·45 g. of a diiodohydroxyazobenzene. This mixture by fractional crystallization from pyridine-ethanol-water afforded two benzoyl derivatives.

The more soluble substance crystallized as light orange crystals from dilute acetic acid, m.p. 138 °C.

Found: C, 41·1; H, 2·2; N, 5·2%.

Calculated for $C_{19}H_{12}O_2N_2I_2$: C, 41·2; H, 2·2; N, 5·1%.

The less soluble derivative formed red-brown granules from acetic acid-ethanol-water, then glacial acetic acid, m.p. 186–187 °C.

Found: C, 35·1; H, 2·4; N, 4·5; I, 54·9%.

The mother liquor did not give a test for iodide ions.

5,3′-Diiodo-2-Hydroxyazobenzene and 4-Iodo-2,6-bis(3′-iodobenzeneazo)-phenol.—Prepared and separated as usual.

Yield of bisazo compound: 11%. It formed near black granular crystals from pyridine-ethanol, m.p. 215–218 °C. (Iodine vapour at 204 °C.)

Found: N, 8·55%.

Calculated for $C_{18}H_{11}ON_4I_3$: N, 8·25%.

Yield of monoazo compound: 19%. This formed orange-brown crystals from acetic acid-ethanol, m.p. 167–168 °C.

Found: N, 6·4%.

Calculated for $C_{12}H_8ON_2I_2$: N, 6·2%.

The benzoyl derivative (Schotten-Baumann), 5,3′-diiodo-2-benzoyloxyazobenzene, formed orange crystals from acetic acid-ethanol-water, m.p. 136 °C., undepressed in m.p. by admixture with above substance $C_{19}H_{12}O_2N_2I_2$.

(i) Action of Sulphuric Acid on

(1) *Potassium 4-Hydroxyazobenzene-4′-sulphonate*

TABLE 14

Expt.	Conditions	Polymer
XI	8·9 g., 50 ml. of 98%, 25 °C., 30 days	0

(2) *3,3′-Dihalogeno-4-hydroxyazobenzenes*

TABLE 15

Expt.	Conditions	Halogen	Recovered Hydroxyazo Compound (%)
	Heated with 98% at 90 °C. for 1 hr.		
XII	0·0465 g., 0·55 ml.	F	87
XIII	0·57 g., 5·94 ml.	Cl	95
XIV	0·76 g., 5·94 ml.	Br	90

(3) *Azoxybenzene*

TABLE 16

Expt.	Conditions	Hydroxyazo Compound	Azo Compound	Polymer	Sulphonic Acids etc.
XV	3·56 g., 50 ml. of 98%, 90 °C., 1 hr.	3%	44%	10%	42–43%

IV. ACKNOWLEDGMENTS

Grateful acknowledgment is here made for the Dunlop Research Scholarship held by one of the authors (P.H.G.), also for the micro- and semi-microanalyses carried out by Misses J. Fildes and B. Naylor, University of Sydney, and Dr. Weiler and Dr. Strauss, Oxford.

V. REFERENCES

ANGELI, A. (1910).—*R.C. Accad. Lincei* [5] **19i** : 793.
BAMBERGER, E., and BAUDISCH, O. (1909).—*Ber. dtsch. chem. Ges.* **42** : 3578.
BAMBERGER, E., and RENAULD, E. (1897).—*Ber. dtsch. chem. Ges.* **30** : 2278.
BRAND, K. (1903).—*J. prakt. Chem.* [2] **67** : 145.
CHEM. FABRIK GRIESHEIM-ELEKTRON (1903).—*Chem. Zbl.* **1903ii** : 320.

Cumming, N. M., and Ferrier, G. S. (1925).—*J. Chem. Soc.* **127** : 2377.
Gabriel, S. (1876).—*Ber. dtsch. chem. Ges.* **9** : 1405.
Gore, P. H., and Hughes, G. K. (1949).—*J. Roy. Soc. N.S.W.* **83** : 266.
Gore, P. H., and Hughes, G. K. (1950).—*Aust. J. Sci. Res.* A **3** : 136.
Heumann, K. (1872).—*Ber. dtsch. chem. Ges.* **5** : 913.
Hodgson, H. H., and Kershaw, A. (1928).—*J. Chem. Soc.* **1928** : 2703.
Jacobsen, P., Heber, E., Henrich, F., and Schwarz, C. (1895).—*Liebigs Ann.* **287** : 184.
Jacobsen, P., and Piepenbrinck, P. (1894).—*Ber. dtsch. chem. Ges.* **27** : 2703.
Knipscheer, H. M. (1903).—*Rec. Trav. Chim. Pays-Bas.* **22** : 1.
Lachman, A. (1902).—*J. Amer. Chem. Soc.* **24** : 1178.
Laubenheimer, A. (1875).—*Ber. dtsch. chem. Ges.* **8** : 1625.
Lukashevish, V. O., and Khurdyumova, T. N. (1949).—*Chem. Abstr.* **43** : 3800.
Meldola, R., and Andrews, E. R. (1896).—*J. Chem. Soc.* **69** : 13.
Neumann, G. S. (1887).—*Liebigs Ann.* **241** : 71.
Nicolet, B. H. (1921).—*J. Amer. Chem. Soc.* **43** : 2081.
Parsons, T., and Bailar, J. C. (1936).—*J. Amer. Chem. Soc.* **58** : 268.
Richter, M. M. (1900).—" Lexicon der Kohlenstoffrerbingungen." Vol. 1. p. 1246. (Verlag von Leopold Voss : Hamburg and Leipzig.)
Schiemann, G., and Pillarsky, R. (1929).—*Ber. dtsch. chem. Ges.* **62** : 3041.
Schultz, G. (1884).—*Ber. dtsch. chem. Ges.* **17** : 464.
Schwarzenbach, G., and Michaelis, L. (1938).—*J. Amer. Chem. Soc.* **60** : 1667.
Vorländer, D., and Meyer, F. (1902).—*Liebigs Ann.* **320** : 129.
Zechmeister, L., and Rom, P. (1929).—*Liebigs Ann.* **468** : 129.

LIQUID–VAPOUR EQUILIBRIA

II. THE SYSTEM BENZENE–*n*-HEPTANE

By I. Brown* and A. H. Ewald*

[*Manuscript received January* 25, 1951]

Summary

Liquid-vapour equilibrium data for the system benzene–*n*-heptane have been measured at intervals over the whole composition range at 60 and 80 °C. and for equimolar mixtures at 50, 70, and 85 °C. The changes in the thermodynamic functions on mixing, both at constant pressure and at constant volume, have been calculated. The values of these changes on mixing at constant volume have been compared with values predicted from component properties. The conditions for azeotrope formation in this system have been defined.

I. Introduction

The liquid-vapour equilibrium concentrations for the system benzene–*n*-heptane were measured as part of an investigation of methods for predicting such data from the properties of the components of binary mixtures. This system was chosen as typical of non-polar systems having a large difference between the internal pressures of the components. The thermodynamic functions for mixing at constant pressure were calculated from the equilibrium data. In order to correct these functions to constant volume mixing conditions it was necessary to determine the coefficient of expansion of the mixture.

II. Experimental

(*a*) *Purification of Materials*

(i) *Benzene.*—Analytical reagent quality benzene was shaken with two portions of concentrated sulphuric acid, washed with sodium carbonate solution and with water, and dried over anhydrous calcium chloride. It was then fractionally crystallized four times and the product was fractionally distilled at a reflux ratio of 70 to 1 through a 1·2 m. column packed with 1·6 mm. Dixon packing ; head and tail fractions were rejected. The physical properties of the purified benzene are shown in Table 1.

(ii) n-*Heptane.*—*n*-Heptane (C.F.R. knock-rating quality) was fractionally distilled in the same column at total reflux. After the removal of three 60 ml. head fractions the column head temperature remained constant within ±0·01 °C. The distillate was then allowed to accumulate in a 500 ml. head trap and was removed after 24 hours operation at total reflux ; three batches of distillate were removed in this way. The physical properties of the bulked product are shown in Table 1.

---

* Division of Industrial Chemistry, C.S.I.R.O., Melbourne.

(b) *Vapour Pressures of the Components*

The vapour pressures of the purified components were measured at 10 degree intervals from 40 to 90 °C. in the apparatus previously described (Brown and Ewald 1950) and the boiling points were measured at 760 mm. Hg (Std.). The Antoine equation*

$$\log P = A - B/(C+t), \qquad \text{(1)}$$

was fitted to the vapour pressure data by the method of Willingham and Rossini (1945) and the constants were found to be :

	Benzene	*n*-Heptane
A	6·91279	6·88686
B	1214·57	1258·27
C	221·166	215·701

Vapour pressure values calculated from equation (1) were found to be 0·08 per cent. higher for benzene and 0·06 per cent. higher for *n*-heptane than those calculated in the same manner by Rossini *et al.* (1947).

Table 1

PHYSICAL PROPERTIES OF COMPONENTS

Physical Property	Benzene		*n*-Heptane	
	This Work	Rossini (1947)	This Work	Rossini
B.p./760 (°C.) ..	$80·07 \pm 0·01$	80·099	98·40	98·427
F.p. (°C.) ..	$+5·48 \pm 0·02$	+5·533	—	—
d_4^{25}	0·87359	0·87370	0·67956	0·67951
N_D^{25}	1·4979	1·49792	1·3851	1·38511

(c) *Analytical Method*

The compositions of liquid and vapour samples were determined from density measurements as described previously for the system carbon tetrachloride-*cyclo*hexane (Brown and Ewald 1950). Thirteen mixtures of benzene–*n*-heptane of known composition were used to determine the constants in equation (2) which expresses the change in specific volume on mixing as a function of composition.

$$\Delta = x_1x_2[0·027243 + 0·005472(x_1-x_2) + 0·004976(x_1-x_2)^2 + 0·007620(x_1-x_2)^3]. \qquad \text{(2)}$$

In Table 2 are shown the mole fractions of benzene, the refractive indices and densities of the synthetic mixtures used for the density-composition calibration, as well as experimental values of Δ and the difference between the latter and values calculated from equation (2). The last column shows the per cent. change in molar volume on mixing at 25 °C.

---

* A table of nomenclature is included at the end of this paper.

TABLE 2

BENZENE-n-HEPTANE PROPERTIES OF MIXTURES OF KNOWN COMPOSITION

x_1	N_D^{25}	d_4^{25}	$\Delta_{\text{exp.}}$	$(\Delta_{\text{exp.}}-\Delta_{\text{calc.}})\times 10^4$	$\frac{100\,V_{25}^M}{\overset{\circ}{V}_{25}}$
0·94305	1·4863	0·85459	0·00239	+1·8	0·205
0·89286	1·4771	0·83920	0·00371	+0·5	0·313
0·87747	1·4746	0·83479	0·00396	−0·7	0·332
0·82371	1·4656	0·81984	0·00502	−0·6	0·413
0·77316	1·4579	0·80677	0·00578	−0·0	0·468
0·77128	1·4578	0·80634	0·00574	−0·7	0·465
0·65798	1·4420	0·78010	0·00671	+0·2	0·526
0·55918	1·4306	0·76002	0·00695	+0·6	0·531
0·43647	1·4179	0·73803	0·00654	−0·0	0·485
0·32594	1·4080	0·72078	0·00556	+0·5	0·403
0·21794	1·3995	0·70564	0·00402	−1·4	0·284
0·16094	1·3951	0·69827	0·00315	−0·2	0·220
0·08211	1·3901	0·68872	0·00176	+1·3	0·122

(*d*) *Equilibrium Measurements*

The apparatus and technique used for the liquid-vapour equilibrium measurements were identical with those used for the system carbon tetrachloride-*cyclo*hexane (Brown and Ewald 1950). Fourteen determinations were made at intervals over the whole composition range at 60 °C. and fifteen at 80 °C. Single determinations were also made at 50, 70, and 85 °C. for approximately equimolar mixtures. The equilibrium still was operated at a vapour throughout equivalent to approximately 15 ml. of condensate per minute and a period of at least three hours was allowed in each determination for the system to approach equilibrium.

(*e*) *The Volume Change on Mixing as a Function of Temperature*

Calculation of the excess thermodynamic functions for mixing at constant volume requires a knowledge of the volume change on mixing, V_t^M.

Values of V_t^M were computed at the required temperatures from measured values of the volume change on mixing at 25 °C., V_{25}^M, using equation (3).

$$V_t^M=K_t\,.\,V_{25}^M=K_t(x_1M_1+x_2M_2)\Delta. \qquad \ldots\ldots\ldots\ldots (3)$$

This involves the assumption that the ratio $V_t^M/V_{25}^M=K_t$ is a constant for all compositions (cf. Wood and Brusie 1943).

Values of K_t were computed at the various temperature levels from equation (4).

$$K_t=\frac{[x_1V_{1(25)}+x_2V_{2(25)}+\Delta(x_1M_1+x_2M_2)]\,[1+\varkappa(t-25)]-[x_1V_{1(t)}+x_2V_{2(t)}]}{(x_1M_1+x_2M_2)\Delta}. \qquad \ldots\ldots\ldots\ldots (4)$$

These were found to be as follows :

t (°C.) ..	25	50	60	70	80
K_t	1·0	0·756	0·720	0·830	0·931

Values of the molar volumes of the components, $V_{1(t)}$ and $V_{2(t)}$ were calculated from the density data given in I.C.T. (1928, p. 28) and by Egloff (1939) respectively and values of Δ were obtained from equation (2). The average coefficients of expansion, $\varkappa$, of a 49·7 mole per cent. benzene mixture were measured over the required temperature range in a Pyrex dilatometer with a bulb of 7 ml. capacity and a 30 cm. capillary stem of 1·88 mm. diameter. The dilatometer was calibrated with pure benzene over the range of 25 to 75 °C. During the calibration and determination it was immersed in a kerosene bath controlled to ±0·01 °C. and the meniscus height above a datum mark was measured with a cathetometer to ±0·05 mm. Values of the coefficients of expansion of this mixture over ranges from 25 to t °C. are shown in Table 3.

TABLE 3

AVERAGE COEFFICIENTS OF EXPANSION OF MIXTURE

Temp. (°C.)	$\varkappa$
44·94	0·001281
49·87	0·001285
59·78	0·001307
70·03	0·001341
75·00	0·001364
79·96	0·001385

III. RESULTS

(*a*) *Equilibrium Data*

Tables 4 and 5 show the experimental values of the equilibrium concentrations and total vapour pressures at temperatures of 60 and 80 °C. together with the relative volatilities calculated by equation (5).

$$\alpha = y_1x_2/y_2x_1. \tag{5}$$

The excess chemical potentials and excess free energy of mixing at constant pressure, which are also shown in these tables, are defined by Scatchard, Raymond, and Gilmann (1938) and represent the variations in the values of the chemical potentials and free energy of mixing from those of an ideal solution of the same composition. They were calculated by the use of the following equations :

$$\mu_1^E = RT \ln \gamma_1 = RT \ln \frac{Py_1}{P_1x_1} + f(V_1-\beta_1)(P_1-P), \tag{6}$$

$$\mu_2^E = RT \ln \gamma_2 = RT \ln \frac{Py_2}{P_2x_2} + f(V_2-\beta_2)(P_2-P), \tag{7}$$

$$G_{xp}^E = x_1\mu_1^E + x_2\mu_2^E. \tag{8}$$

The vapour pressure values required were calculated from equation (1). The second virial coefficients for benzene are those computed by Tompa (1948),

Table 4

BENZENE-*n*-HEPTANE. EQUILIBRIUM DATA AND DERIVED FUNCTIONS AT 60 °C.

x_1	y_1	P (mm. Hg)	μ_1^E (cal./mol.)	μ_2^E (cal./mol.)	G^E_{xp} (cal./mol.)	G^E_{xp} (difference)	y_1 (difference)	α
0·01789	0·04218	216·43	181·6	1·9	5·1	+0·5	—0·0063	2·417
0·04544	0·09996	224·86	160·9	4·4	11·5	+0·1	—0·0088	2·333
0·09274	0·19302	239·66	165·6	7·1	21·8	—0·2	—0·0105	2·340
0·18792	0·34165	266·92	146·4	15·4	40·0	—0·1	—0·0066	2·243
0·29336	0·47169	294·35	128·8	24·9	55·4	—0·1	—0·0071	2·150
0·38715	0·56254	315·58	107·1	39·1	65·4	+0·3	—0·0056	2·035
0·48775	0·64567	335·65	85·4	57·4	71·3	+0·2	—0·0030	1·914
0·57844	0·70782	350·83	62·1	87·6	72·8	+0·6	—0·0044	1·766
0·58017	0·70915	351·01	61·7	87·6	72·6	+0·4	—0·0043	1·765
0·68364	0·77608	365·93	39·7	125·8	67·0	—0·6	—0·0031	1·604
0·79118	0·84285	378·15	19·0	188·5	54·4	—0·5	—0·0025	1·410
0·89576	0·91354	387·64	6·0	270·3	33·5	+0·4	—0·0003	1·228
0·94253	0·94947	390·29	2·4	313·3	20·3	+0·5	+0·0003	1·147
0·97900	0·98021	391·51	0·4	361·2	7·9	+0·2	—0·0000	1·073

Table 5

BENZENE-*n*-HEPTANE. EQUILIBRIUM DATA AND DERIVED FUNCTIONS AT 80 °C.

x_1	y_1	P (mm. Hg)	μ_1^E (cal./mol.)	μ_2^E (cal./mol.)	G^E_{xp} (cal./mol.)	G^E_{xp} (difference)	y_1 (difference)	α
0·02113	0·04288	440·14	125·7	2·9	5·5	+0·4	—0·0054	2·076
0·04373	0·08641	452·34	125·8	5·1	10·4	+0·2	—0·0094	2·068
0·04580	0·09008	453·06	123·6	4·9	10·4	—0·3	—0·0096	2·060
0·09287	0·17231	478·77	120·5	11·3	21·5	+0·9	—0·0156	2·033
0·18947	0·32242	527·72	126·5	15·6	36·6	—0·7	—0·0117	2·035
0·28835	0·43994	572·65	105·7	28·2	50·5	+0·1	—0·0118	1·938
0·39089	0·54459	614·60	90·1	39·6	59·3	—0·5	—0·0074	1·863
0·49515	0·63154	651·90	68·2	62·0	65·1	+0·2	—0·0073	1·748
0·58192	0·69303	678·53	47·3	92·9	66·4	+0·9	—0·0104	1·622
0·61584	0·71970	688·11	43·5	97·9	64·4	—0·3	—0·0070	1·602
0·68479	0·76680	706·53	31·4	125·1	61·0	—0·1	—0·0061	1·513
0·79214	0·83614	730·14	12·2	191·5	49·5	+0·0	—0·0071	1·339
0·88321	0·90384	746·65	5·6	236·8	32·6	—0·1	—0·0013	1·243
0·94237	0·94939	754·04	1·3	288·6	17·9	—0·1	—0·0006	1·147
0·97934	0·98097	757·20	0·1	324·9	6·8	—0·1	—0·0002	1·091

while those for *n*-heptane were calculated from the Berthelot equation using the critical constants $T_c = 540$ °K. and $P_c = 26 \cdot 8$ atm. The liquid molar volumes of benzene and *n*-heptane were computed at the required temperatures as in Section II (*e*). Both the second virial coefficients and molar volumes used are shown in Table 6.

Table 6

SECOND VIRIAL COEFFICIENTS AND MOLAR VOLUMES (LITRES)

t (°C.)	Benzene		*n*-Heptane	
	β_1	V_1	β_2	V_2
50	—1·150	0·092	—1·831	0·152
60	—1·090	0·094	—1·716	0·154
70	—1·010	0·095	—1·610	0·157
80	—0·980	0·096	—1·514	0·159
85	—0·928	0·097	—1·469	0·159

It was found that the values of G^E_{xp} shown in Tables 4 and 5 could be expressed as a function of composition by equation (**9**)

$$G^E_{xp} = x_1 x_2 [a + b(x_1 - x_2) + c(x_1 - x_2)^2], \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{9})$$

where a, b, and c have the following values:

$$a = 718 \cdot 231 - 1 \cdot 297\, T,$$
$$b = 228 \cdot 223 - 0 \cdot 5084\, T,$$
$$c = 40 \cdot 667 - 0 \cdot 01713\, T.$$

These were obtained by using the method of least squares to find the best values of the coefficients at 60 and 80 °C. and by assuming these to vary linearly with temperature.

The corresponding equations for the excess chemical potentials are

$$\mu^E_1 = x_2^2 [a - b + x_1(4b - c) + 6c(x_1^2 + \tfrac{1}{6} x_2 - x_1 x_2)], \quad \ldots\ldots\ldots\ldots \quad (\mathbf{10a})$$

$$\mu^E_2 = x_1^2 [a + b + x_2(-4b - c) + 6c(x_2^2 + \tfrac{1}{6} x_1 - x_1 x_2)]. \quad \ldots\ldots\ldots\ldots \quad (\mathbf{10b})$$

In the seventh column of Tables 4 and 5 is shown the difference between values of G^E_{xp} calculated from equations (**6**), (**7**), and (**8**) and those calculated using equation (**9**). In the eighth column is shown the difference between the experimental values of the vapour concentration and those calculated using equation (**10**) together with values of P_1 and P_2 from equation (**1**), of β and V from Table 6, and experimental values of x_1 and P.

Table 7 shows additional equilibrium measurements for approximately equimolar mixtures at 50, 70, and 85 °C. The differences shown in the last column indicate that equation (**9**) is suitable for temperature interpolation; the differences at 50 and 85 °C. which involve a temperature extrapolation are larger than the standard deviations of G^E_{xp} (0·42 and 0·51 cal./g. mol. at 60 and 80 °C. respectively) but are still well within the estimated errors in G^E_{xp} (see Section III (*e*)).

Table 7

BENZENE-*n*-HEPTANE. EQUILIBRIUM DATA AND DERIVED FUNCTIONS AT OTHER TEMPERATURES

Temp. (°C.)	x_1	y_1	P (mm. Hg)	μ^E_1 (cal./mol.)	μ^E_2 (cal./mol.)	G^E_{xp} (cal./mol.)	G^E_{xp} (difference)
85	0·49246	0·62093	757·35	60·1	69·0	64·7	+1·4
80	0·49515	0·63154	651·90	68·2	62·0	65·1	+0·2
70	0·49080	0·63644	472·57	74·9	61·9	68·2	+0·2
60	0·48775	0·64567	335·65	85·4	57·9	71·3	+0·2
50	0·48616	0·65259	232·35	90·7	60·8	75·3	+1·0

(*b*) *The Excess Entropy of Mixing and Heat of Mixing at Constant Pressure*

The excess entropy of mixing may be obtained by differentiation of equation (9) with respect to temperature. Since $G^E_{xp}=H^M_{xp}-TS^E_{xp}$, we find that the temperature independent part of the coefficients of equation (9) represents H^M_{xp} and the temperature dependent part represents $-TS^E_{xp}$. We therefore have the equations

$$S^E_{xp}=x_1x_2[1{\cdot}297+0{\cdot}5084(x_1-x_2)+0{\cdot}01713(x_1-x_2)^2], \quad \ldots\ldots \quad (11)$$

and

$$H^M_{xp}=x_1x_2[718{\cdot}231+228{\cdot}223(x_1-x_2)+40{\cdot}667(x_1-x_2)^2]. \quad \ldots\ldots \quad (12)$$

Values of S^E_{xp} and H^M_{xp} calculated from these equations at even mole fractions are shown in Table 8.

(*c*) *Azeotrope Formation*

To determine the conditions for azeotrope formation for this system the values of the relative volatilities for temperatures of 60 and 80 °C. shown in Tables 4 and 5 were plotted against x_1 and the curves produced to $x_1=1$. Since the curves did not intersect the line $\alpha=1$ no azeotrope is formed even at 60 °C. The change of relative volatility with temperature was found by combining equations (5), (6), and (7) to give

$$RT\ln\alpha=RT\ln(P_1/P_2)+\mu^E_1-\mu^E_2-f[(V_1-\beta_1)(P_1-P)-(V_2-\beta_2)(P_2-P)]. \quad \ldots\ldots \quad (13)$$

As x_1 approaches unity, equation (13) reduces to

$$RT\ln\alpha=RT\ln(P_1/P_2)-\mu^E_2+f(V_2-\beta_2)(P_2-P_1). \quad \ldots\ldots \quad (14)$$

TABLE 8

BENZENE-n-HEPTANE. THERMODYNAMIC FUNCTIONS AT CONSTANT PRESSURE AND CONSTANT VOLUME
(cal./mol. and cal./deg. mol.)

x_1	G^E_{xp} (70 °C.)	S^E_{xp} (70 °C.)	H^M_{xp} (70 °C.)	G^E_{xv} (80 °C.)	G^E_{xv} (60 °C.)	S^E_{xv} (70 °C.)	H^M_{xv} (70 °C.)	$\mathbf{E}^M_{xv}$ (70 °C.)*	$\mathbf{S}^E_{xv}$ (70 °C.)*	$\mathbf{E}^M_{xv}-T\mathbf{S}^E_{xv}$*
0·02	4·97	0·016	10·52	8·62	8·54	—0·004	7·18	4·74	0·001	4·56
0·05	12·01	0·041	25·92	21·32	21·24	—0·004	19·80	11·61	0·010	8·15
0·1	22·71	0·081	50·55	41·63	41·46	—0·008	38·66	22·47	0·020	15·71
0·2	40·54	0·160	95·35	77·54	77·33	—0·011	73·79	41·66	0·036	29·27
0·3	54·00	0·230	133·02	105·63	105·63	—0·000	105·86	57·12	0·049	40·30
0·4	63·29	0·287	161·81	125·11	125·52	+0·021	132·35	68·34	0·058	48·54
0·5	68·27	0·324	179·56	135·83	136·71	+0·044	151·40	74·69	0·062	53·34
0·6	68·45	0·336	183·72	137·49	138·74	+0·062	159·49	75·40	0·061	54·36
0·7	63·07	0·316	171·37	128·95	130·31	+0·068	152·93	69·58	0·056	50·43
0·8	50·85	0·257	139·17	107·45	108·60	+0·058	127·89	56·07	0·044	41·00
0·9	30·45	0·154	83·42	67·57	68·23	+0·033	79·23	33·46	0·026	24·51
0·95	16·61	0·084	45·44	37·95	38·30	+0·017	44·06	18·22	0·014	13·52
0·98	7·00	0·035	19·11	16·30	16·44	+0·007	18·78	7·66	0·006	5·64

* Symbols in bold type refer to values predicted from component properties.

This and equation (**10b**) were used to calculate the following limiting values of α as x_1 approaches unity.

t (°C.) ..	80	60	50	40	30
α ..	1·059	1·033	1·018	1·003	0·988.

Graphical interpolation of these values shows that $\alpha=1$ at 38 °C., so that an azeotrope should form at high benzene concentrations below this temperature. The argument assumes the validity of equation (**10**) down to 38 °C.

No definite experimental evidence for azeotrope formation with this system has been published. Marschner and Cropper (1946) report an azeotrope which contains 99·5 mole per cent. benzene and boils at 80·1 °C. at 760 mm. Hg, but state that there is doubt as to its existence. Stage and Schultze (1944) found no azeotrope at 760 mm. Hg. Sieg (1950), who measured constant pressure equilibria from 760 to 300 mm. Hg (approximately 80 to 53 °C.) found no evidence of azeotrope formation.

(*d*) *Values of the Excess Thermodynamic Functions for Mixing at Constant Volume*

The thermodynamic functions calculated here from experimental data refer to mixing at constant pressure, while those predicted from the properties of the component liquids refer to constant volume conditions. In order to compare the two sets of functions it was therefore necessary to convert the former to constant volume conditions by use of the following relations of which equation (**15**) was given by Scatchard (1949).

$$G_{xv}^{E}=G_{xp}^{E}+\frac{V_{t}^{M}}{\omega}, \quad \ldots\ldots \quad (\mathbf{15})$$

$$S_{xv}^{E}=-[G_{xv}^{E}\,(80\ ^\circ\mathrm{C.})-G_{xv}^{E}\,(60\ ^\circ\mathrm{C.})]/20, \quad \ldots\ldots \quad (\mathbf{16})$$

$$H_{xv}^{M}=G_{xv}^{E}+TS_{xv}^{E}. \quad \ldots\ldots \quad (\mathbf{17})$$

Values of these excess functions for mixing at constant volume calculated using these equations are shown in Table 8. Values of V_t^M were computed from equation (**3**) and values of the isothermal compressibility of the mixture, ω, obtained from the compressibilities of the pure components assuming linear variation with mole fraction. The effect on G_{xv}^E and S_{xv}^E of using mole fractions rather than volume fractions is not significant. The compressibility data for benzene were obtained from I.C.T. (1928, p. 38) and those for *n*-heptane by graphical interpolation between the value at 99·5 °C. of Amagat given in I.C.T. (1928, p. 37) and values at 20, 30, 40, and 50 °C. calculated from the adiabatic compressibilities of Freyer, Hubbard, and Andrews (1929). The values used were as follows:

t (°C.)	ω Benzene (atm.$^{-1}$)	ω *n*-Heptane (atm.$^{-1}$)
80	156×10^{-6}	231×10^{-6}
60	128×10^{-6}	197×10^{-6}

(e) *Errors in the Derived Thermodynamic Functions*

Before the significance of the values of these functions could be determined it was necessary to evaluate the errors in them due to errors in the measured quantities from which they were calculated. For temperature, pressure, and composition the errors were ±0·01 °C., ±0·05 mm. Hg, and ±0·00025 mole fraction respectively, while in the derived functions for an equimolar mixture at an average temperature of 70 °C. they were:

Function	Absolute Error (±)
μ^E	2·0 cal./mol.
G^E_{xp}	2·0 cal./mol.
G^E_{vx}	3·9 cal./mol.
S^E_{xp}	0·20 cal./deg. mol.
S^E_{xv}	0·23 cal./deg. mol.

Values of μ^E_1 from equation (**6**) were less than, and those of μ^E_2 from equation (**7**) were greater than the corresponding values calculated from equation (**10**). Thus the experimental values show some thermodynamic inconsistency, which is probably due to low values of the vapour concentration (see Tables 4 and 5). This is more noticeable at low mole fractions of benzene where the relative volatility is higher than in benzene-rich solutions. It is probably due to evaporation of small amounts of benzene-rich mixture from the vapour trap during the sampling period.

(f) *Comparison of Thermodynamic Functions Calculated from Experimental Data with those Predicted from the Properties of the Component Liquids*

To make this comparison the following relation was employed:

$$G^E_{xp}=E^M_{xv}-TS^E_{xv}+\frac{V^\circ}{\omega}[\ln(1+V^M/V^\circ)-V^M/V^\circ], \quad \ldots\ldots \quad (\mathbf{18})$$

which follows from the relations given by Scatchard (1949). For this system the last term is of the order of 0·001 cal./mol. and can be neglected. The remaining terms on the right-hand side can be predicted from the properties of the components following the methods of Hildebrand and Scatchard (1936), Flory (1942), and Huggins (1942).

Values of E^M_{xv} can be calculated for regular solutions from the relation*

$$\mathbf{E}^M_{xv}=(x_1V_1+x_2V_2)v_1v_2D^2, \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{19})$$

where

$$D=(\Pi_1^{0\cdot5}-\Pi_2^{0\cdot5}) \text{ and } \Pi_1=\Lambda_1/V_1.$$

Values of $\mathbf{E}^M_{xv}$ at 70 °C. were calculated using values of Π_1 and Π_2 obtained from the heats of vaporization using the relation $\Lambda=L-RT$. Values of L at 70 °C. were determined from the vapour pressures using equation (**1**) and the relation

$$L_1=RT^2\frac{d\ln P_1}{dT}\left[1+\frac{P_1}{RT}(\beta_1-V_1)\right]. \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{20})$$

---

* Symbols in bold type indicate quantities predicted from component properties.

The values found were 7500 and 8116 cal./g. mol. for benzene and *n*-heptane respectively. Values of L were also calculated from the heats of vaporization at the normal boiling points given by Rossini *et al.* (1947) using the equation of Watson (1943) for extrapolation. They were 7486 and 8038 cal./g. mol. respectively. Using the values calculated from equation (**20**) it was found that $\Pi_1=71\cdot87$ cal./ml., $\Pi_2=47\cdot44$ cal./ml., and $D^2=2\cdot528$. The excess entropy changes on mixing (shown in Table 8) which are due only to differences in the molar volumes of the components were obtained by the use of equation (**21**) which was derived by Flory and Huggins for polymer solutions.

$$\mathbf{S}_{xv}^{E}=R[x_1 \ln V_1+x_2 \ln V_2-\ln (x_1V_1+x_2V_2)]. \quad \ldots\ldots (\mathbf{21})$$

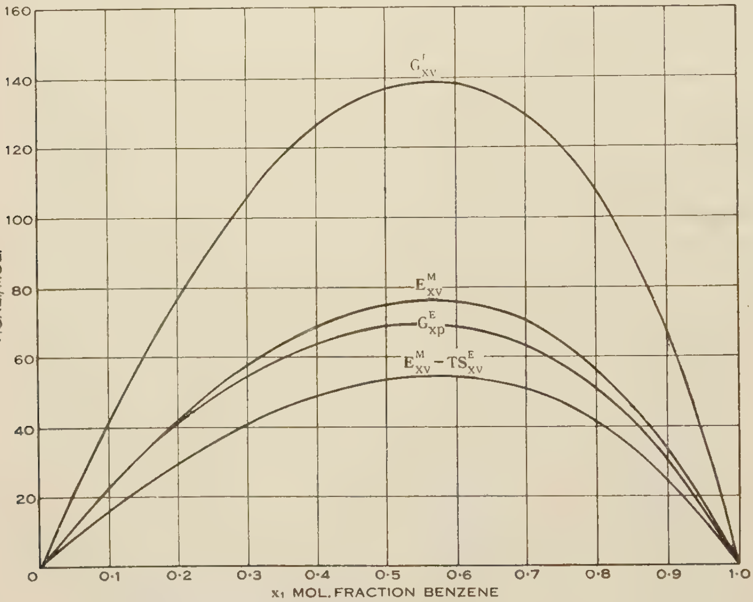


Fig. 1 —Benzene-*n*-heptane. Thermodynamic functions at 70 °C.

In comparing predicted values with values of G_{xp}^{E} calculated from the equilibrium data it is necessary to consider the errors that could occur in the predicted values of $\mathbf{E}_{xv}^{M}$. The fractional error in the internal energy of mixing determined from equation (**19**) is given by

$$\frac{\delta\mathbf{E}_{xv}^{M}}{\mathbf{E}_{xv}^{M}}=\frac{3\delta V}{V}+\frac{\Pi_1^{0\cdot5}+\Pi_2^{0\cdot5}}{\Pi_1^{0\cdot5}-\Pi_2^{0\cdot5}}\left[\frac{\delta V}{V}+\frac{\delta\Lambda}{\Lambda}\right], \quad \ldots\ldots\ldots\ldots\ldots (\mathbf{22})$$

where $\delta V/V$ and $\delta\Lambda/\Lambda$ are the fractional errors in the molar volumes and energies of vaporization respectively. Thus the error in $\mathbf{E}_{xv}^{M}$ increases rapidly as the difference in the internal pressures becomes smaller. Using conservative estimates of the fractional errors in Λ and V of $0\cdot004$ and $0\cdot005$ respectively we find for the system benzene–*n*-heptane at 70 °C. that the error in $\mathbf{E}_{xv}^{M}$ at $x_1=0\cdot5$ is $7\cdot7$ cal./mol. ($10\cdot2$ per cent.). For the system carbon tetrachloride-benzene where $D=0\cdot475$ the error in $\mathbf{E}^{M}_{xv}$ would be as high as 36·6 per cent. Thus, it is not surprising that large differences are often found between experimental values of G^{E}_{xp} for regular systems and those calculated using equation (**19**). Conversely the apparent good agreement between internal pressures calculated from experimental values of G^{E}_{xp} and those calculated from calorimetric values of the heat of vaporization may be misleading (see, for example, Scatchard, Wood, and Mochel 1939*a*, 1939*b*).

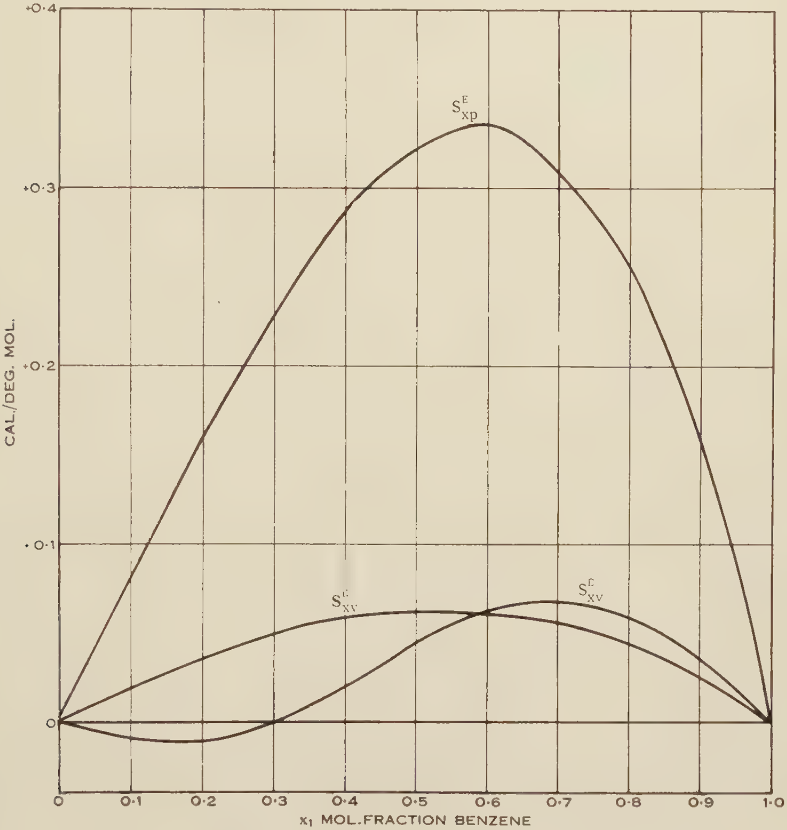


Fig. 2.—Benzene-*n*-heptane. Excess entropy at 70 °C.

A comparison of values of G^{E}_{xp} at 70 °C. determined from the experimental data with those predicted from the properties of the component liquids at the same temperature is shown in Figure 1.

IV. Discussion

The significance of the excess free energies of mixing at constant pressure and at constant volume is not affected by their errors which are of the order of 3 per cent. By comparison of values of G^{E}_{xv} and G^{E}_{xp} shown in Figure 1 it can be seen that $G^{E}_{xv}\approx 2G^{E}_{xp}$ or $G^{E}_{xp}\approx V^{M}_{t}/\omega$ (compare equation (**15**)), which is in agreement with the relation derived by Scatchard (1937).

For the excess entropy of mixing at constant pressure the error is about 0·20 cal./deg. mol. or equivalent to 70 per cent. of its maximum value. In spite of this large error, the experimental values prove S^E_{xp} to be positive. The error in S^E_{xv}, however, is 0·23 cal./deg. mol. and is equivalent to four times its maximum value; S^E_{xv} is therefore not significantly different from zero. The negative values of S^E_{xv} at low benzene concentrations shown in Figure 2 could be due to the relatively large experimental error at low benzene concentration. The excess entropy of this system is thus due mostly to the change of volume on mixing; but positive values of S^E_{xv}, if they should prove significant, could be due to the breakdown of order in the benzene as is suggested by Scatchard (1939*b*) for the system benzene-*cyclo*hexane.

It is only for regular solutions that the function E^M_{xv} is equal to the excess free energy of mixing. For this system, $\mathbf{E}^M_{xv}$ as calculated from equation (**19**) is equal to G^E_{xp} within the error of $\mathbf{E}^M_{xv}$, in agreement with the finding that $S^E_{xv} \approx 0$. The values of $\mathbf{S}^E_{xv}$ calculated by the Flory-Huggins equation (**21**) and shown in Figure 2 agree with experimental values as to sign and order of magnitude over most of the concentration range. This comparison does not provide a check whether the Flory-Huggins equation is applicable to systems of this type, because of the large experimental error in S^E_{xv}.

V. CONCLUSIONS

(1) An azeotrope in the system benzene–*n*-heptane should be found at high benzene concentrations at temperatures below 38 °C.

(2) The Hildebrand-Scatchard expression for the internal energy of mixing at constant volume gives a reasonable agreement with experimental values of the excess free energy of mixing at constant pressure for this system, but preliminary study of other systems shows the method to be unsuitable for precise prediction of excess free energies of near-regular solutions. The study of additional mixtures will be necessary to define the range of systems for which this method is useful.

(3) A wider study will also be required to test the combined use of the Hildebrand-Scatchard and Flory-Huggins theories in predicting the behaviour of non-regular systems. In future work it is proposed to measure calorimetrically the heats of mixing and also to measure the volume changes on mixing at the same temperature used for the liquid-vapour equilibrium measurements in order to obtain more accurate values of the excess entropy of mixing at constant volume.

VI. ACKNOWLEDGMENTS

The authors thank Mr. M. H. Taubert, chief chemist, Vacuum Oil Co., Melbourne, and Mr. J. Poppleton, assistant chief technical officer, Shell Co. of Australia Ltd., for gifts of *n*-heptane; also Mr. J. A. Barker and Dr. M. E. Winfield, Division of Industrial Chemistry, C.S.I.R.O., for valuable suggestions during the preparation of the manuscript.

VII. Table of Nomenclature

A, B, C = Constants in the Antoine equation (**1**),
D = $(\Pi_1^{0\cdot5}-\Pi_2^{0\cdot5})$,
E = Internal energy,
$\mathbf{E}$ = Internal energy predicted from component properties,
G = Free energy (Gibbs),
H = Enthalpy,
K_t = V_t^M/V_{25}^M,
L = Latent heat of vaporization,
M = Molecular weight,
N_D^{25} = Refractive index at 25 °C. for sodium D line,
P = Pressure,
P_1, P_2 = Vapour pressures of components 1 and 2,
R = Gas constant,
S = Entropy,
$\mathbf{S}$ = Entropy predicted from component properties,
T, t = Temperature °K., °C.,
V = Molar volume,
V° = $x_1V_1+x_2V_2$,
d_4^{25} = Density at 25 °C. referred to water at 4 °C.,
f = Constant for converting from mm. Hg-litres to calories $=0\cdot0318657$,
v = Volume fraction,
x, y = Mole fraction in liquid and vapour respectively,
α = Relative volatility,
$\varkappa$ = Average coefficient of expansion,
β = Second virial coefficient in equation of state,
ω = Isothermal compressibility,
γ = Activity coefficient,
$\delta(\ )$ = Absolute error in ( ),
Δ = Change in specific volume on mixing at 25 °C.,
μ = Chemical potential,
Λ = Energy of vaporization,
Π = Internal pressure,

Superscripts

E = Change in excess of that for an ideal solution,
M = Change on mixing,

Subscripts

1, 2 refer to the more and less volatile component respectively,
x indicates a function of composition,
p indicates mixing at constant pressure,
v indicates mixing at constant volume,
t indicates a function of temperature,
c refers to critical state.

VIII. REFERENCES

BROWN, I., and EWALD, A. H. (1950).—*Aust. J. Sci. Res.* A **3**: 306.

EGLOFF, G. (1939).—"Physical Constants of Hydrocarbons." Vol. 1. p. 40. (Reinhold: New York.)

FLORY, P. J. (1942).—*J. Chem. Phys.* **10**: 51.

FREYER, E. B., HUBBARD, J. C., and ANDREWS, D. H. (1929).—*J. Amer. Chem. Soc.* **51**: 759.

HILDEBRAND, J. H. (1936).—"Solubility of Non-Electrolytes." (Reinhold: New York.)

HUGGINS, M. L. (1942).—*Ann. N.Y. Acad. Sci.* **43**: 1.

INTERNATIONAL CRITICAL TABLES (1928).—Vol. 3. (McGraw Hill: New York.)

MARSCHNER, R. F., and CROPPER, W. P. (1946).—*Industr. Engng. Chem.* **38**: 262.

ROSSINI, F. D., *et al.* (1947).—Selected values of properties of hydrocarbons. Circ. U.S. Nat. Bur. Stand. (Dep. Comm.) C.461.

SCATCHARD, G. (1937).—*Trans. Faraday Soc.* **83**: 160.

SCATCHARD, G. (1949).—*Chem. Rev.* **44**: 7.

SCATCHARD, G., RAYMOND, C. L., and GILMANN, H. H. (1938).—*J. Amer. Chem. Soc.* **60**: 1275, 1278.

SCATCHARD, G., WOOD, S. E., and MOCHEL, J. M. (1939*a*).—*J. Amer. Chem. Soc.* **61**: 3206.

SCATCHARD, G., WOOD, S. E., and MOCHEL, J. M. (1939*b*).—*J. Phys. Chem.* **43**: 119.

SIEG, L. (1950).—*Chemie Ing. Tech.* **22**: 322.

STAGE, H., and SCHULTZE, G. R. (1944).—*Oel. u. Kohle* **40**: 66, 68, 90.

TOMPA, H. (1948).—*J. Chem. Phys.* **16**: 292.

WATSON, K. M. (1943).—*Industr. Engng. Chem.* **35**: 398.

WILLINGHAM, C. B., *et al.* (1945).—*J. Res. Nat. Bur. Stand.* **35**: 219.

WOOD, S. E., and BRUSIE, J. P. (1943).—*J. Amer. Chem. Soc.* **65**: 1891.

THE DIELECTRIC PROPERTIES OF SYMMETRICAL LONG-CHAIN SECONDARY ALCOHOLS IN THE SOLID STATE

By R. J. Meakins* and R. A. Sack†

[*Manuscript received October 6, 1950*]

Summary

Symmetrical long-chain secondary alcohols in the solid state show very high dielectric loss at audio and radio frequencies. This can be explained by the presence of chains of hydroxyl groups linked by hydrogen bonding and capable of reversing their direction. Further evidence of hydrogen bonding is provided by a study of the melting points of the secondary alcohols and related compounds.

The amount of dielectric loss depends markedly on the manner of formation of the solid, being smallest for samples formed by recrystallization from solvents at low temperatures and largest for specimens obtained by slow cooling from the melt. The alcohols of molecular chain-lengths of 13, 15, 17, and 19 carbon atoms show a considerable decrease of absorption on storing at room temperature. For alcohols of between 23 and 43 carbon atoms the loss is rather smaller with a peak at higher frequencies, but remains more constant in time.

These results are interpreted in terms of competing influences of van der Waals forces and hydrogen bonds during crystal formation; the former, which lead to a structure unsuitable for the formation of hydrogen-bond chains, are predominant at low temperatures, but become more rapidly neutralized by thermal motion, especially for the shorter molecules. The high temperature modification of the lower homologues is unstable at room temperature, and a molecular diffusion process causes the bond chains to break.

Dilute systems of secondary alcohols with hydrocarbons or paraffin wax of similar molecular chain-length show very small dielectric loss suggesting a solid solution in which bond chains cannot be formed; if the paraffin molecules are appreciably longer, the absorption is large and decreases on storing, presumably owing to the presence of a pure alcohol phase.

I. INTRODUCTION

Previous investigations of the dielectric properties of organic long-chain compounds in the solid state have been carried out mainly with substances in which the polar group and the molecular chain form a comparatively rigid unit. In these materials dipole orientation involves rotation of the entire molecule. Evidence of this is obtained from measurements with pure esters (Meakins 1949) and with solid systems of esters (Sillars 1938) and ketones (Meakins 1949) in paraffin wax and hydrocarbons, the relaxation time in each case increasing regularly with the chain-length of the polar molecules.

In many organic compounds dipole orientation can take place by rotation of the polar group about a single chemical bond. Thus, methyl palmitate shows

---

* Division of Industrial Chemistry, seconded to the Division of Electrotechnology, C.S.I.R.O., University Grounds, Sydney.

† Division of Electrotechnology, C.S.I.R.O., University Grounds, Sydney.

two regions of dielectric absorption (Dryden and Jackson 1948 ; Jackson 1949), one at radio frequencies, due to the orientation of complete molecules, and another in the microwave region, corresponding probably to independent orientation of the methoxy or carbomethoxy groups (Dryden and Welsh, personal communication).

In secondary alcohols the hydroxyl groups can rotate about their C–O bonds and some preliminary results with symmetrical normal long-chain alcohols have already been reported (Meakins and Sack 1949). That the dielectric absorption is not due to orientation of entire molecules was indicated by the result that it was not displaced towards lower frequencies with increasing molecular weight. The values of the dielectric loss factor were, however, very large compared with those of esters of similar chain-lengths, although the dipole moments of alcohols and esters are not very different. Moreover, the frequency of maximum absorption was lower by several decades than would be expected for the rotation or orientation of individual hydroxyl groups. The results could be explained, however, by the presence of hydrogen-bonded chains of hydroxyl groups capable of reversal (see Fig. 1) ; calculations have shown that both the amount of dielectric absorption and the effective relaxation time should increase with the length of the bond chains. The theory that the losses in pure secondary alcohols are due to such a mechanism was supported by the absence of absorption in certain dilute solid systems of secondary alcohols with paraffins in which the hydrogen-bond chains would be broken by dispersal of the polar molecules.

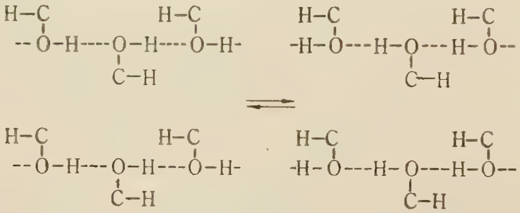


Fig. 1.—Diagram showing reversal mechanism of H-bond chain.

The present paper describes a more detailed investigation of the dielectric properties of symmetrical long-chain secondary alcohols in the solid state, both pure and in mixtures with hydrocarbons and paraffin wax. The effects of varying the conditions of formation of the crystals were examined, particular interest being paid to the different properties of the pure alcohols in the recrystallized form obtained on precipitation from solvents or in the " melted " form in which the material was first heated to the liquid state and solidified on cooling.

II. Apparatus and Procedure

The compounds or mixtures for measurement were prepared in the form of disks, 5 cm. in diameter and about $0 \cdot 2$ cm. in thickness. The method previously described of preparing " melted " specimens by casting and machining (Meakins

1949) was abandoned in favour of the following technique. About 3·0–3·5 g. of material was weighed into a small glass beaker, melted in an oven, and allowed to solidify by transferring it to the laboratory atmosphere. The solid was then carefully powdered and pressed into disks under a total load of about 5 tons weight. Sheets of aluminium foil were placed in the mould to facilitate the removal of the pressed disk. The dielectric properties of melted samples prepared by either method are similar, but the pressing technique is not only quicker and requires less material, but is also applicable to the preparation of recrystallized samples from the crystalline powder, thus enabling a direct comparison between the two modifications.

The measurements of loss angle and dielectric constant at frequencies between 100 c/s. and 50 kc/s. were made by a substitution bridge method, using a Wagner earth technique in conjunction with guard-ring electrodes.

In the frequency range 50 kc/s. to 50 Mc/s. measurements of the loss angle and dielectric constant were made by a resonance method as previously used (Meakins 1949).

The electrical measurements were made within 2 hours of preparation of the disks except where otherwise stated. The temperature during measurement was 20±2 °C.

The preparation of the materials used is described in Appendix I where their melting points are also given.

III. Results and Discussion

(a) *Melting Points*

The distinctive nature of the hydroxyl group as compared with other polar units is evident from a study of the melting points of different series of symmetrically substituted long-chain compounds as shown in Figure 2. In a long paraffin molecule, the replacement of a hydrogen atom by a polar group usually leads to a reduction in melting point. Presumably the introduction of the substituent causes an expansion of the crystal lattice, and the resulting decrease in the van der Waals forces leads to a net reduction in the lattice energy in spite of the additional attraction due to the polar groups. For the symmetrical secondary alcohols, however, the melting points are higher than those of the parent hydrocarbons, which indicates the presence of marked attractive forces connected with the hydroxyl groups ; these forces are most simply accounted for by hydrogen bonding.

(b) *Dielectric Properties of the Pure Alcohols*

(i) *Comparison of Melted and Recrystallized Forms*.—Considerable differences were found in the dielectric absorption of crystals prepared by cooling from the liquid state and by recrystallization from solution. Figure 3 shows the results for 10-nonadecanol ($C_{19}H_{40}O$) in which the maximum dielectric loss factor ε'' for specimens obtained from the melt is more than 100 times larger than that of crystals precipitated from ethanol. Material recrystallized from light petroleum gave almost identical results with that from ethanol. 10-Nonadecanol is highly soluble in these solvents, and freezing is required to cause precipitation. It is

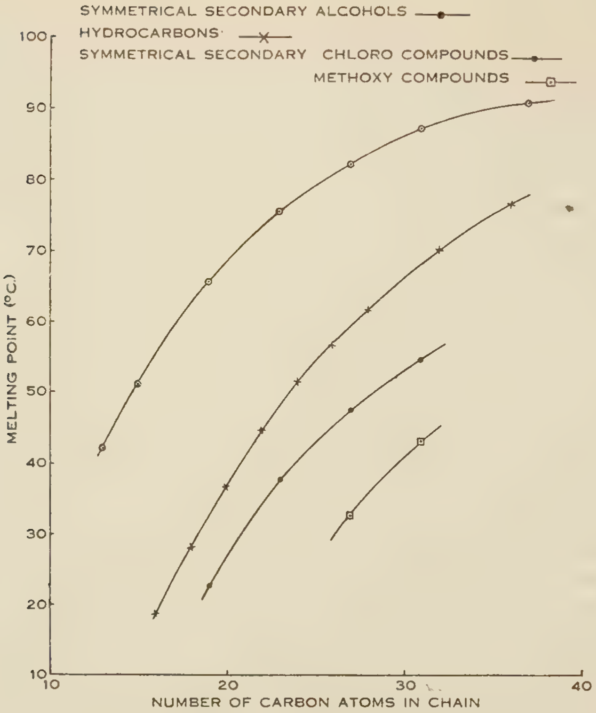


Fig. 2.—Melting points of various long-chain compounds.

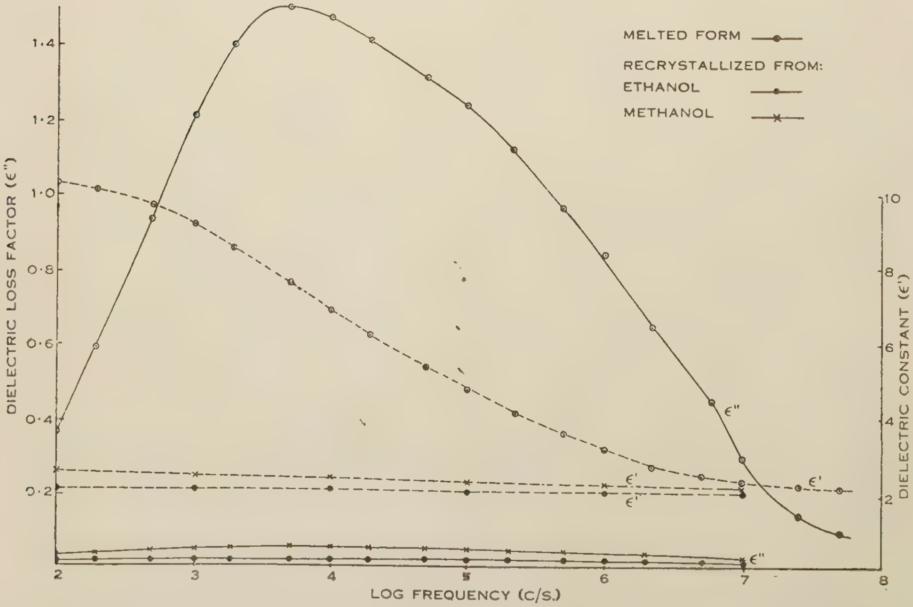


Fig. 3.—Dielectric absorption of melted and recrystallized forms of 10-nonadecanol ($C_{19}H_{40}O$).

less soluble in methanol and crystallizes at a higher temperature. A sample obtained by precipitation from methanol showed a loss factor about three times as large as those from ethanol or light petroleum.

These results suggest that the temperature of formation of the crystals may be an important factor in determining their dielectric properties at room temperature. Further evidence for this assumption is provided by the results of experiments with 14-heptacosanol ($C_{27}H_{56}O$). This is comparatively insoluble, and at a concentration of 1 : 10 in ethanol precipitates at about 40 °C. Figure 4 shows that crystals prepared in this way give a maximum dielectric loss factor which is smaller than that of the melted form in a ratio of the order of 1 : 6 only.

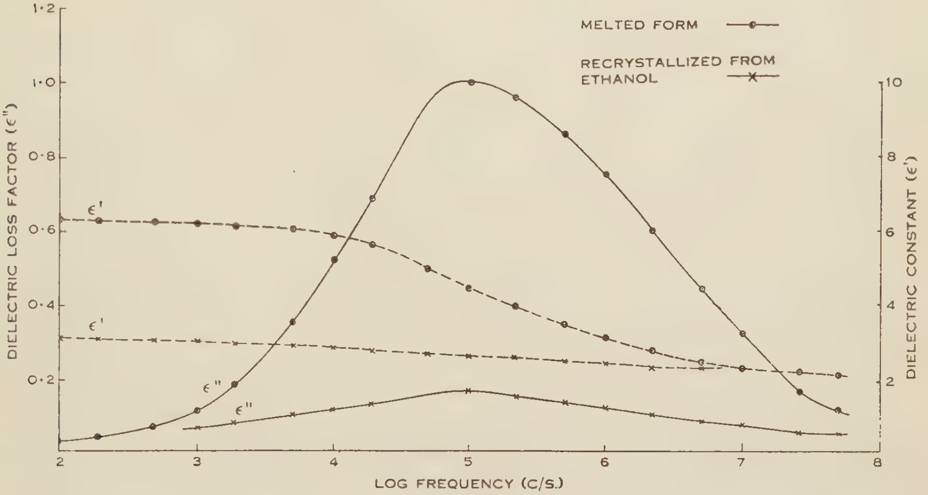


Fig. 4.—Dielectric absorption of melted and recrystallized forms of 14-heptacosanol ($C_{27}H_{56}O$).

With both 10-nonadecanol and 14-heptacosanol the higher loss factors ε'' for the melted samples are accompanied by a larger increase in the dielectric constant ε' towards lower frequencies. This is in agreement with the theorem derived by Gross (1941).

The above results show that the formation of long hydrogen-bonded chains of hydroxyl groups in secondary alcohols is favoured by high temperatures of crystallization. The total lattice energy of the melted and recrystallized forms, on the other hand, is apparently very similar as there is no significant difference in melting point.

(ii) *Effect of Rate of Formation of Crystals from the Melt.*—Even for specimens that are obtained by solidification from the liquid state the method of preparation greatly affects the dielectric properties. Figure 5 shows the ε' and ε'' curves for two samples of 16-hentriacontanol ($C_{31}H_{64}O$) prepared from the melt; one sample was allowed to crystallize in the usual way, surrounded by air at room temperature, and the other kept, during solidification, in an incubator at 80 °C.

(m.p. 87 °C.). It is seen that the more slowly crystallized sample gives considerably higher dielectric absorption, with a displacement of the maximum towards lower frequencies. This indicates the presence of shorter hydrogen-bond chains in the rapidly cooled specimen ; this difference may arise because of the greater speed of solidification leading to a larger number of defects in the lattice ; an alternative explanation is that we are dealing with a temperature effect similar

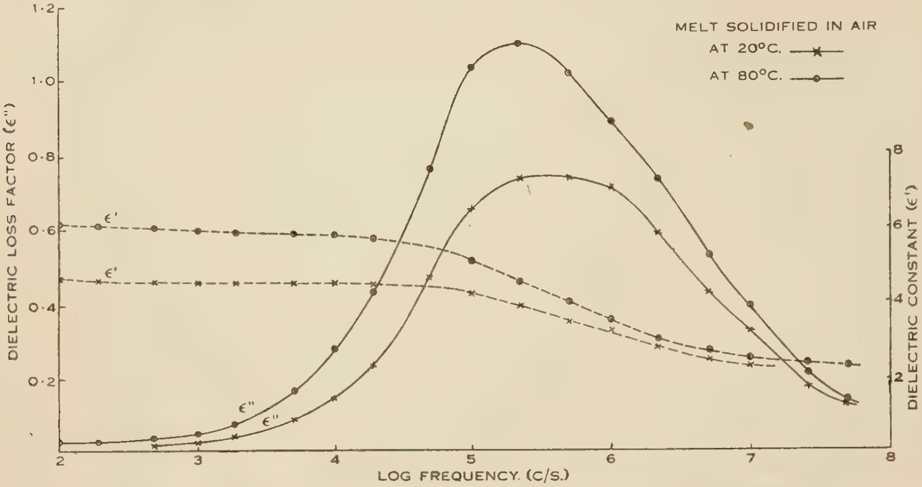


Fig. 5.—Effect of rate of solidification on dielectric absorption of 16-hentriacontanol ($C_{31}H_{64}O$).

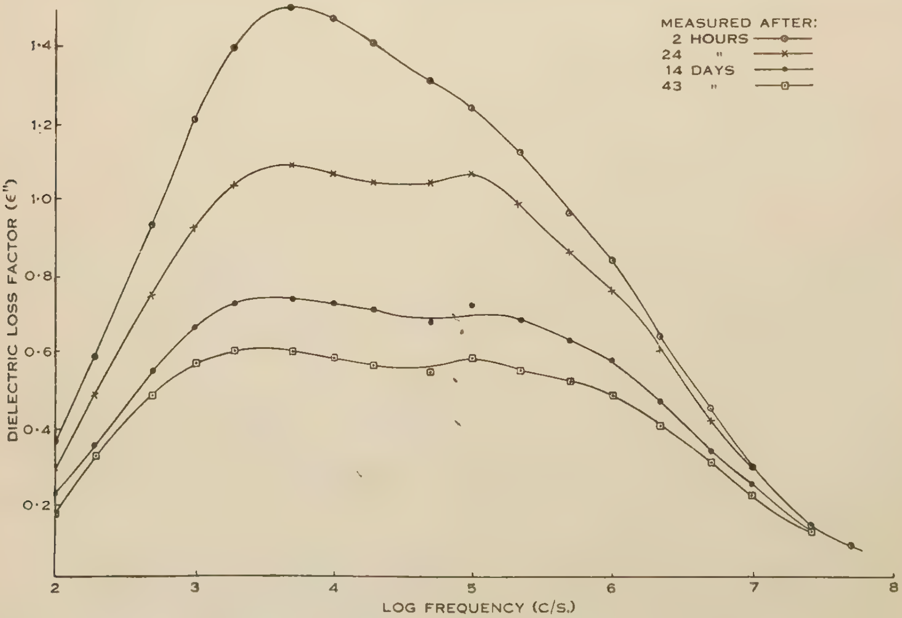


Fig. 6.—Change in dielectric absorption of 10-nonadecanol ($C_{19}H_{40}O$) on storing.

to that described in the preceding paragraph, the solids being actually formed at different temperatures owing to varying degrees of super-cooling.

(iii) *Reduction of Absorption on Storage.*—The dependence of strong dielectric absorption on a high temperature of formation of the specimen raises the question whether the highly absorbing state is actually stable at room temperature or whether a molecular rearrangement will take place leading to a reduction in the absorption. The latter effect was found to occur with the lower homologues of the series. Figure 6 shows the loss factor/frequency curve for a sample of 10-nonadecanol immediately after preparation and after various periods of storage at room temperature. The decrease in ε'' is, of course, accompanied by a decrease in the low frequency values of ε'. The alcohols of smaller molecular weight, 7-tridecanol ($C_{13}H_{28}O$), 8-pentadecanol ($C_{15}H_{32}O$), and 9-heptadecanol ($C_{17}H_{36}O$), show similar changes during storage at room temperature. With 12-tricosanol ($C_{23}H_{48}O$) the change, though still appreciable, is much smaller, whereas with 14-heptacosanol ($C_{27}H_{56}O$) and the secondary alcohols of longer chain-length negligible changes are observed during periods of up to 80 days.

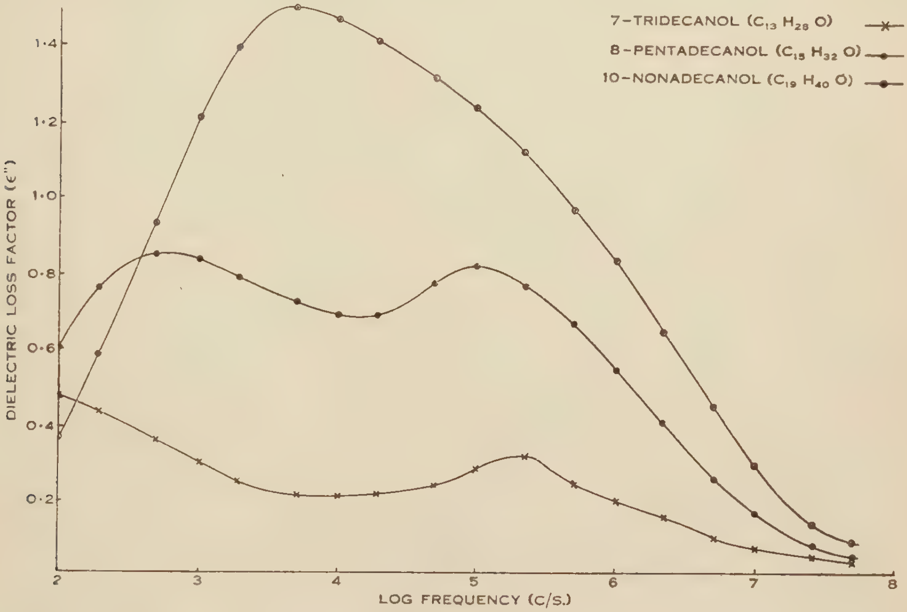


Fig. 7.—Dielectric absorption of crystalline secondary alcohols (melted form).

These results suggest that in the secondary alcohols the observed changes in dielectric properties on storage are due to a rearrangement involving molecular diffusion, which, for the higher members of the series, is slowed up or entirely suppressed by the increase in activation energy.

(iv) *Variation of Dielectric Absorption with Molecular Weight.*—The variation of the dielectric loss factor/frequency curves with the chain-length of the symmetrical secondary alcohols is shown in Figures 7 and 8. In the first group

(Fig. 7) comprising alcohols of chain-lengths of up to 19 carbon atoms, 7-tridecanol (m.p. 42 °C.) and 8-pentadecanol (m.p. 51 °C.) show a rather low dielectric absorption in comparison with 10-nonadecanol (m.p. 66 °C.) ; this may be due to their lower temperatures of solidification or to a more rapid molecular rearrangement during the period between preparation and measurement of the samples.

In passing from 10-nonadecanol to 12-tricosanol (Fig. 8) there is a considerable decrease in the dielectric loss especially in the lower frequency range, indicating the presence of shorter hydrogen-bond chains. With further increase in molecular weight, the absorption continues to be displaced to higher frequencies, but to a much smaller extent. There is also a further decrease in the amount of absorption which, no doubt, is partially due to the increasing dilution of the polar groups.

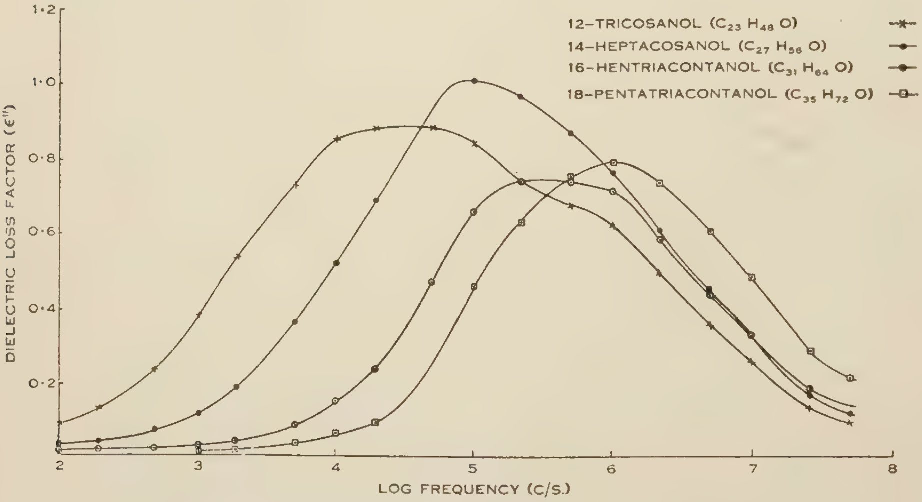


Fig. 8.—Dielectric absorption of crystalline secondary alcohols (melted form).

(*c*) *Solid Systems with Paraffins*

(i) *Systems with Pure Hydrocarbons.*—As previously reported (Meakins and Sack 1949), 0·1 molal systems of 14-heptacosanol ($C_{27}H_{56}O$) in *n*-octacosane ($C_{28}H_{58}$) and of 16-hentriacontanol ($C_{31}H_{64}O$) in *n*-dotriacontane ($C_{32}H_{66}$) gave negligible dielectric absorption over the wide range of frequencies studied. In these systems the chain-length of the solvent exceeds that of the solute by one carbon atom. Figure 9 shows, however, that for systems of 12-tricosanol ($C_{23}H_{48}O$) in the same concentration with three different hydrocarbons the absorption increases as the chain-length of the solvent becomes increasingly longer than that of the solute ; the system with *n*-triacontane ($C_{30}H_{62}$) shows a loss factor corresponding approximately to that of the pure alcohol in proportion to its concentration in the mixture. The same applied to a 0·1m system of 10-nonadecanol ($C_{19}H_{40}O$) in *n*-octacosane ($C_{28}H_{58}$). These results suggest that for mixtures in which the length of the hydrocarbon molecule considerably exceeds that of the alcohol, H-bonding is still present ; presumably a proportion, at least, of the pure alcohol precipitates during cooling. As the chain-length of the paraffin approaches that of the solute, most of the alcohol molecules appear to be dispersed in the form of a solid solution, and the hydrogen-bond chains are either not formed or are of insufficient length to cause an appreciable absorption in the frequency range investigated.

It is of interest to note that in the latter case there is no indication of an absorption of the type found in solid solutions of similar concentrations of ketones or esters in paraffins (Meakins 1949), which is associated with rotations of the polar molecules as a whole. As mentioned before (Meakins and Sack 1949) this suggests that the large size of the singly-bonded oxygen atom prevents the molecules from performing rotational transitions.

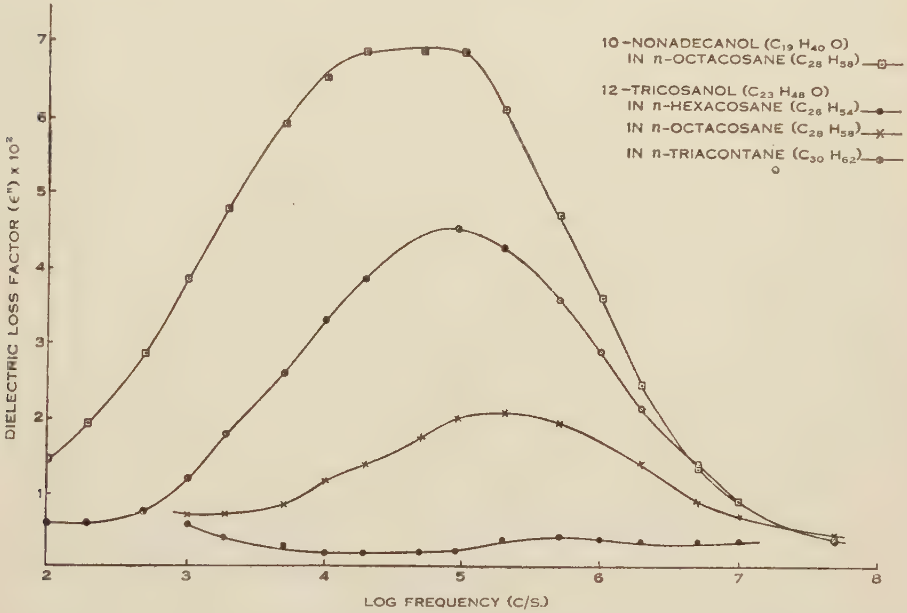


Fig. 9.—Dielectric absorption of 0·1m systems of secondary alcohols with hydrocarbons.

As expected from previous results (Section III (*b*) (iii)), systems in which a portion of the alcohol ingredient separates as a pure phase show a decrease in dielectric absorption during storage at 20 °C. Figure 10 shows this effect for the two systems of Figure 9 which yielded the highest losses. With the 10-nonadecanol in *n*-octacosane, the diminution corresponds to that of the pure alcohol, but with 12-tricosanol in *n*-triacontane the reduction in absorption is considerably larger than would be expected during the same period from the results on pure 12-tricosanol. This may be due in part to a diffusion of secondary alcohol molecules into the crystal lattice, or else to a more rapid rearrangement in the expanded structure which would be expected to result from the mixing.

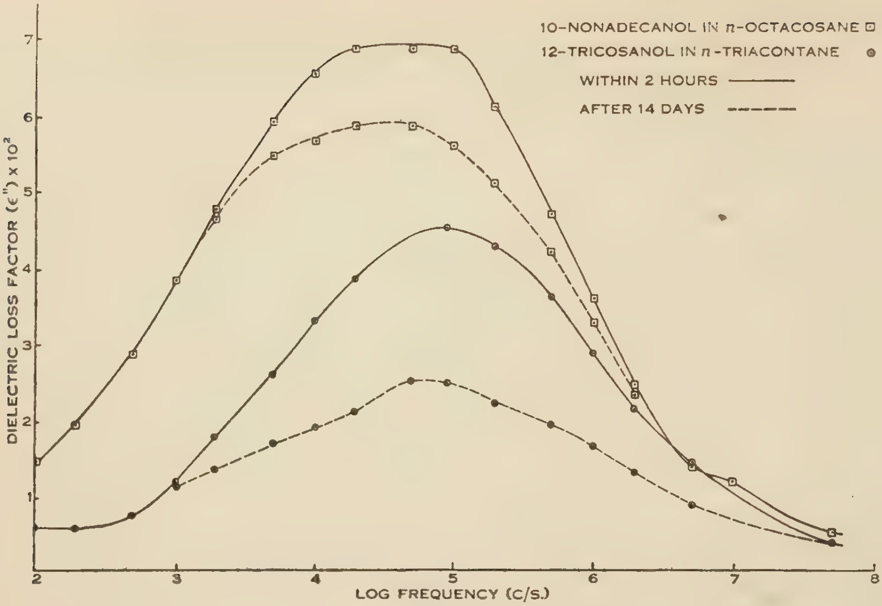


Fig. 10.—Change in dielectric absorption of 0·1m systems with hydrocarbons on storing.

(ii) *Systems with Paraffin Wax.*—Systems of secondary alcohols with paraffin wax (0·1 molal) show similar dielectric absorption (see Fig. 11) to those in which pure paraffins are used as solvents. The wax used in these experiments had a melting point of 62–65 °C., corresponding to an average chain-length of about

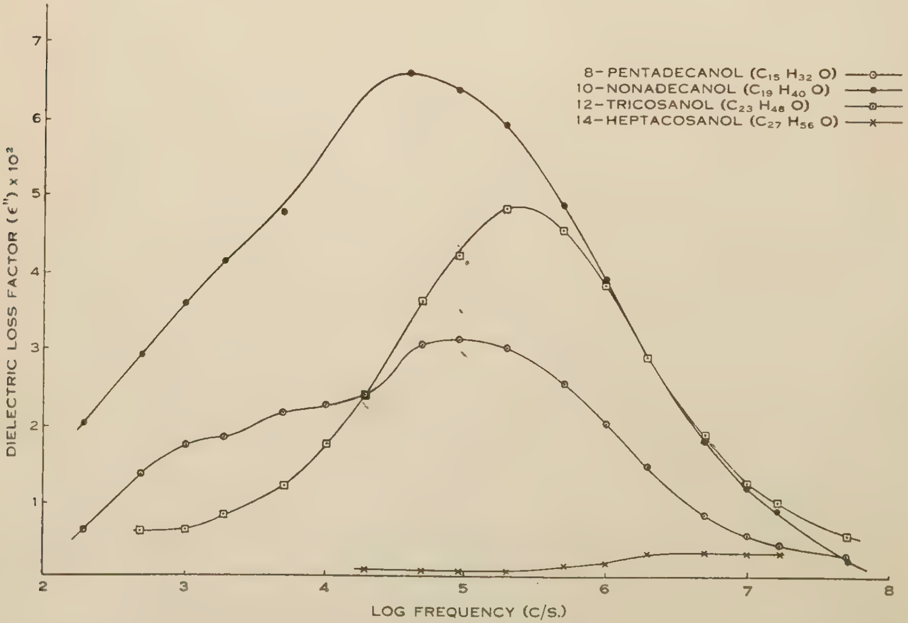


Fig. 11.—Dielectric absorption of 0·1m systems of secondary alcohols with paraffin wax.

29 carbon atoms. It is seen that for 14-heptacosanol, with a similar chain-length, the loss is small, though not quite negligible, whereas for shorter alcohols the absorption corresponds again to a nearly complete separation into a pure alcohol phase and the wax. For these systems the decrease in dielectric absorption during storage is more pronounced than with either the pure alcohols or their systems with pure hydrocarbons. Figure 12 gives the results for the 10-nonadecanol system which shows a very rapid initial decrease and becomes practically stable after 14 days. A similar rate of decrease and proportionate change was observed for 12-tricosanol in paraffin wax. The results obtained 15 months after preparation apply to different samples prepared by the former method of melting and casting (Meakins 1949), but both gave good agreement with the results obtained on the later specimens a fortnight after preparation. It appears that the more open structure of the paraffin wax facilitates diffusion or rearrangement of the alcohol molecules.

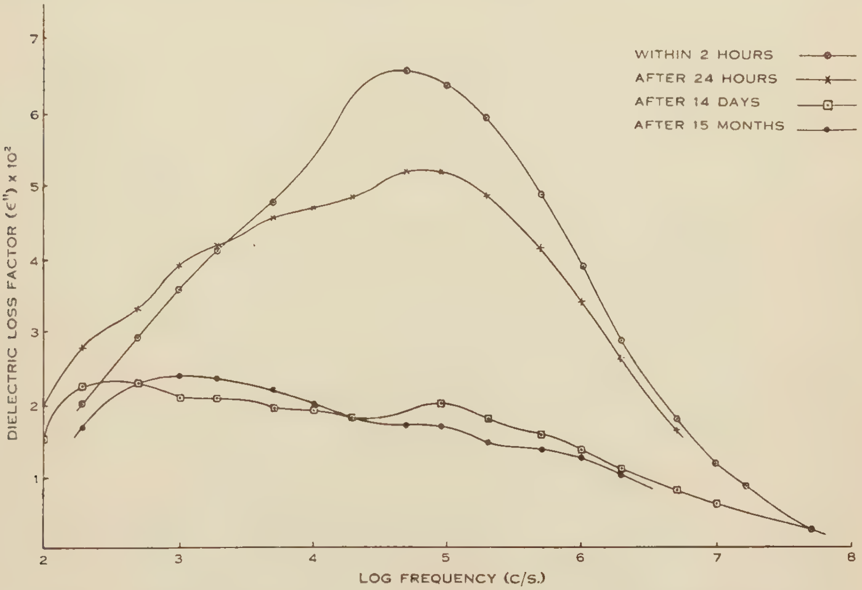


Fig. 12.—Change in dielectric absorption of 0·1m 10-nonadecanol in paraffin wax on storage.

IV. General Discussion

The results presented in the preliminary note (Meakins and Sack 1949) and in this paper have led to the theory that the high dielectric absorption shown by symmetrical long-chain secondary alcohols at audio and radio frequencies is due to the presence of hydrogen-bonded chains of hydroxyl groups. It has already been suggested (Meakins and Sack 1949) that one possible mechanism for effecting chain reversal would be a rotation of the individual groups about their C–O bond. All the dipoles in the chain need not be parallel, but any strong dependence of the frequency of maximum absorption on the number of hydroxyl groups constituting the chain can only be obtained if arrangements, in which two consecutive protons point towards each other, such as

. . O–H . . . O–H H–O . . . H–O . .

are impossible, e.g. for steric reasons ; this means that there can be, at the most, one pair of antiparallel neighbours of the type

. . H–O . . . H–O O–H . . . O–H . .

in each chain. Such a discontinuity can wander along the chain or pass right out of it by means of reversals of individual dipoles, the reversal of the whole chain shown in Figure 1 corresponding to a transfer of the discontinuity from one end to the other.

An alternative possibility is that each hydrogen atom has two positions of equilibrium closer to one or other of its adjoining oxygen atoms and is capable of transitions between these across the intervening energy barrier, but in such a way that no two protons are attached to the same oxygen atom at the same time ; this would require some sort of activation of the protons at the ends of the chain. The mathematical formulation is similar for either of the proposed mechanisms and, as will be shown in a later paper, leads to a loss factor/frequency relation indistinguishable from a Debye curve with a total absorption proportional to the cube and a relaxation time proportional to the square of the number of hydrogen bonds in the chain. The number of such bonds must be of the order of hundreds, at least, to account for the magnitude and frequency of the dielectric absorption observed in the long-chain secondary alcohols. The wide frequency range of the absorption suggests that the lengths of the bond chains are not uniform, but vary within wide limits.

It is unlikely that both of the proposed mechanisms are present together since, as has been shown by Stearn and Eyring (1937) in the case of water (cf. also Baker and Yager 1942), their coexistence would enable the protons to travel along the chains under the action of a constant field and thus lead to an appreciable D.C. conductivity, whereas measurements by the present authors with 14-heptacosanol showed a conductivity of an order not exceeding that of paraffin wax.

Theories similar to the above have been previously proposed to account for the properties of other substances showing abnormal dielectric behaviour. Thus a reversal mechanism of the first type, but extending over three hydrogen bonds only has been suggested to account for the ferro-electricity of Rochelle salt (Beevers and Hughes 1941). The alternative mechanism, involving proton transfer, has been inferred for potassium dihydrogen phosphate (Slater 1941) and ice (Pauling 1935), the former showing spontaneous polarization and the latter a static dielectric constant of over 80.

An outstanding feature of the solid secondary alcohols is the magnitude of their absorption and dielectric constant in comparison with the number of hydroxyl groups per unit volume. The alcohol, 10-nonadecanol, which shows the highest loss has a dielectric constant ε' of 10·4 at 50 c/s., whereas for liquid n-amyl alcohol with more than three times the concentration of dipoles, ε'

rises only to 15·8 at low frequencies (cf., e.g. Kirkwood 1946). It appears that the enhancement of the polar effect is due to the crystal structure which forces the hydrogen-bond chains to develop in straight lines, whereas in liquids, and to some extent in potassium dihydrogen phosphate and ice, the chains are kinked with a statistical distribution.

A more detailed description of the role played by the hydroxyl groups in the solid secondary alcohols cannot be given at the present stage as their crystal structure is not known. Certain tentative conclusions may be drawn from the known lattice structure of normal long-chain hydrocarbons (Müller 1928; Bunn 1939). In these the C–C bonds within each molecule form a plane zigzag, and the molecules are stacked side by side in planes. In the normal state, which most of the paraffins take up at room temperature, the molecular chains are perpendicular to the planes, and the base of the unit cell forms a rectangle of dimensions $a \approx 7 \cdot 5$ Å, $b \approx 5$ Å, which contains the projections of two molecules. Measurements made on long-chain ketones gave similar results (Oldham and Ubbelohde 1939). Some of the hydrocarbons, especially when obtained by recrystallization, have an alternative structure in which the axes of the molecules are tilted from the normal to the base plane. With increasing temperature most paraffins pass through one or several phase transitions, and near the melting point their lattice is hexagonal (Müller 1930, 1932).

The only X-ray measurements on long-chain secondary alcohols reported to date have been determinations of the long spacing (c-axis) of the recrystallized form (Piper *et al.* 1931) which indicate that in some cases the molecular chains are normal to the base plane and in others tilted. With the melted form the explanation put forward here for the observed dielectric properties is not compatible with the assumption of a normal lattice with the same dimensions as that of the corresponding hydrocarbon, for a closer inspection of this structure (Müller 1928; Bunn 1939) shows that for steric reasons hydrogen-bond chains could not develop in random directions or parallel to the short axis without considerable distortion of the molecules. Also, if the bond chains stretched in the direction of the a-axis, the distance between two successive oxygen atoms would slightly exceed $\frac{1}{2}a = 3 \cdot 8$ Å, whereas the O–O distance in hydrogen bonds is usually found to be between $2 \cdot 5$ and $2 \cdot 8$ Å (Pauling 1944).

The results of Section III (b) (i), showing an increase in dielectric absorption of secondary alcohols with increasing temperature of crystallization, suggest that the arrangement of the molecules is determined by the competing influences during crystal growth of van der Waals intermolecular forces and of hydrogen bonding. The former are predominant at low temperatures and presumably lead to a lattice structure similar to one of the paraffin modifications and thus unsuitable for the formation of long hydrogen-bond chains. The van der Waals forces, being distributed all along the length of the molecules, are more strongly counteracted by an increase in thermal motion than the localized short-range hydrogen bonds. The influence of the latter on the crystal structure thus becomes more important the higher the temperature of formation. With samples prepared by cooling from the melt the temperature of solidification is higher than the melting point of the corresponding paraffins, and hence the van der Waals forces are relatively insignificant. Consequently, the hydrogen bonds, which are present even in the liquid, impress on the crystal a structure of their own which enables the bond chains to be formed freely. This high temperature modification appears, however, to be unstable at room temperature, at any rate for the lower homologues, and a gradual rearrangement takes place in which the bonds are broken and the absorption is correspondingly reduced (Section III (*b*) (iii)).

The influence of the van der Waals forces can be further weakened if the crystalline order is reduced by mixing together alcohol molecules of different lengths. To test this assumption, a preliminary set of measurements was carried out on a sample obtained by solidification of a mixture of liquid 12-tricosanol and 14-heptacosanol in equal weights. This sample showed a higher dielectric absorption than either of its components in the pure state, indicating the enhanced influence of the hydrogen bonds in the mixture.

As the length of the molecules increases, the relative importance of the hydroxyl groups decreases ; this causes a smaller difference in melting points of the longer secondary alcohols and their corresponding hydrocarbons and a weaker development of bond chains which can be inferred from the lower dielectric loss (Section III (*b*) (iv)). The rather discontinuous reduction of the absorption as a function of the molecular weight and the disappearance of its decrease during storage for alcohols containing 23 or more carbon atoms suggest a qualitative change in crystal structure compared with the lower members of the series ; but this hypothesis will have to be confirmed by other experiments.

It is of interest to note that very high dielectric constants and loss factors have been observed with long-chain primary alcohols in an unstable phase just below the melting point, in particular in the experiments of Hoffman and Smyth (1949). These authors interpret their results in terms of molecular rotations and state that there is no indication of a rotation of the hydroxyl groups about the C–O bond. The mechanism suggested in this paper involving straight chains of hydrogen bonds offers an alternative explanation which naturally accounts for the increase of the dielectric constant on solidification without the additional postulate of a Maxwell-Wagner effect due to inclusions of a second phase in the solid.

V. ACKNOWLEDGMENT

The authors wish to thank Mr. E. Filby, Division of Electrotechnology, C.S.I.R.O., for carrying out the measurements at frequencies between 100 c/s. and 50 kc/s.

VI. REFERENCES

BAKER, W. O., and YAGER, W. A. (1942).—*J. Amer. Chem. Soc.* **64** : 2171.
BEEVERS, C. A., and HUGHES, W. (1941).—*Proc. Roy. Soc.* A **177** : 251.
BUNN, C. W. (1939).—*Trans. Faraday Soc.* **35** : 482.
DRYDEN, J. S., and JACKSON, W. (1948).—*Nature* **162** : 656.
GROSS, B. (1941).—*Phys. Rev.* **59** : 748.
HOFFMAN, J. D., and SMYTH, C. P. (1949).—*J. Amer. Chem. Soc.* **71** : 431.
JACKSON, W. (1949).—*Nature* **164** : 486.

KIRKWOOD, J. G. (1946).—*Trans. Faraday Soc.* **42A**: 7.
MEAKINS, R. J. (1949).—*Aust. J. Sci. Res.* A **2**: 405.
MEAKINS, R. J., and SACK, R. A. (1949).—*Nature* **164**: 798.
MÜLLER, A. (1928).—*Proc. Roy. Soc.* A **120**: 437.
MÜLLER, A. (1930).—*Proc. Roy. Soc.* A **127**: 417.
MÜLLER, A. (1932).—*Proc. Roy. Soc.* A **138**: 514.
OLDHAM, J. W. H., and UBBELOHDE, A. R. (1939).—*Trans. Faraday Soc.* **35**: 328.
PAULING, L. (1935).—*J. Amer. Chem. Soc.* **57**: 2680.
PAULING, L. (1944).—"Nature of the Chemical Bond." (Cornell University Press: Ithaca, N.Y.)
PIPER, S. H., CHIBNALL, A. C., HOPKINS, S. J., POLLARD, A., SMITH, J. A. B., and WILLIAMS, E. F. (1931).—*Biochem. J.* **25**: 2072.
SILLARS, R. W. (1938).—*Proc. Roy. Soc.* A **169**: 66.
SLATER, J. C. (1941).—*J. Chem. Phys.* **9**: 16.
STEARN, A. E., and EYRING, H. (1937).—*J. Chem. Phys.* **5**: 113.

APPENDIX I

Preparation and Melting Points of Materials

The secondary alcohols were prepared by reduction of the corresponding ketones with aluminium *iso*propoxide and *iso*propyl alcohol, the yields being, in all cases, greater than 90 per cent.

The preparation and properties of the hydrocarbons have been previously described (Meakins 1949).

TABLE 1

Compound	Melting Point (uncorr.) (°C.)
7-Tridecanol ($C_{13}H_{27}OH$)	41·8–42·4
8-Pentadecanol ($C_{15}H_{31}OH$)	51·0–51·3
10-Nonadecanol ($C_{19}H_{39}OH$)	65·9–66·1
12-Tricosanol ($C_{23}H_{47}OH$)	75·2–75·6
14-Heptacosanol ($C_{27}H_{55}OH$)	81·8–82·4
16-Hentriacontanol ($C_{31}H_{63}OH$)	86·4–87·4
18-Pentatriacontanol ($C_{35}H_{71}OH$)	89·8–90·8
10-*Chlorononadecane* ($C_{19}H_{39}Cl$)	22·0–23·5
12-*Chlorotricosane* ($C_{23}H_{47}Cl$)	37·0–37·8
14-*Chloroheptacosane* ($C_{27}H_{55}Cl$)	47·3–47·8
16-*Chlorohentriacontane* ($C_{31}H_{63}Cl$)	53·8–55·4
14-*Methoxyheptacosane* ($C_{27}H_{55}OCH_3$)	32·2–32·5
16-*Methoxyhentriacontane* ($C_{31}H_{63}OCH_3$)	42·5–43·5

The chloro and methoxy derivatives of the alcohols, referred to in Figure 2, were preliminary small scale preparations only. The chloro compounds were prepared by treating the appropriate secondary alcohol with phosphorus pentachloride. The former, suspended in light petroleum, was shaken at room temperature with an equal weight of phosphorus pentachloride until solution occurred, and then set aside for a further 2 hours. The excess phosphorus pentachloride was decomposed by the addition of ethanol, the solvent evaporated in a vacuum, and the residue recrystallized at least twice from ethanol. Under these conditions there appeared to be little tendency for unsaturated by-products to form, and the yields of the chloro compounds were in the region of 90 per cent.

The methoxy compounds were also prepared from the secondary alcohols, the latter being first converted to the sodium derivative by stirring with sodium for about 12 hours at a temperature of 180 °C. A small excess of methyl iodide was then added and the mixture heated under reflux for 100 hours. The product was decomposed with steam and water and then shaken with light petroleum. The latter solution was then dried and chromatographed on alumina, elution with light petroleum giving the required methoxy compound, free of secondary alcohol. The yields were 45–50 per cent.

The melting points (uncorrected) of the secondary alcohols and chloro and methoxy compounds investigated in this programme are set out in Table 1.

CORRIGENDA

VOLUME 3, NUMBER 1, PAGES 91–104

"*Self-Diffusion in Tin Crystals.*" By P. J. Fensham

In the original paper, some arithmetic errors were made in calculating diffusion coefficients from the diffusion-penetration curves, with consequent errors in the quoted results. The new values for Tables 3 and 4 are given below, together with an amended Figure 3.

TABLE 3

PRINCIPAL DIFFUSION COEFFICIENTS

Temperature (t °C.)	Diffusion Coefficient (cm.^{2} sec.$^{-1}$) D_c	D_a
180·5	$11 \cdot 1 \times 10^{-11}$	$5 \cdot 27 \times 10^{-11}$
197·0	$14 \cdot 8 \times 10^{-11}$	$6 \cdot 01 \times 10^{-11}$
210·7	$21 \cdot 3 \times 10^{-11}$	$9 \cdot 30 \times 10^{-11}$
223·1	$26 \cdot 5 \times 10^{-11}$	$9 \cdot 29 \times 10^{-11}$

TABLE 4

ARRHENIUS CONSTANTS FOR TIN

Constant	Parallel to a-Axis	c-Axis
A (cm.^{2} sec.$^{-1}$)	$9 \cdot 2 \pm 2 \times 10^{-8}$	$3 \cdot 6 \pm 1 \times 10^{-6}$
Q (cal./g. atom)	$6 \cdot 7 \pm 1 \times 10^3$	$9 \cdot 4 \pm 0 \cdot 5 \times 10^3$

These alterations do not affect the discussion in any way, except that the temperature at which D_c should equal D_a is now 88 °C. In particular, it can be seen that the high value of Q is still associated with the high value of A as has recently been found also in the non-metallic case of diffusion of Ra^{++} into albite and micropertite.*

* ROSENQUIST, I. TH. (1949).—*Acta Chem. Scand.* **3**: 569.

Furthermore, some discussion has taken place on the validity of the solution of the diffusion equation that has been used. In this connection, it should be noted that under the conditions of electrodeposition used, the effective (dense) active layer involved in the diffusion had a thickness of the order of only 0·001 mm. rather than the values (0·02–0·03 mm.) given on page 94.

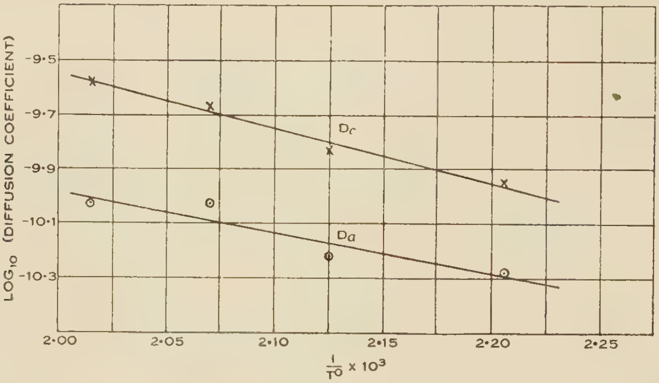


Fig. 3.—Logarithms of D_c and D_a as functions of temperature.

The author would like to thank Mr. J. F. Nicholas for drawing his attention to the errors and for collaboration in the re-calculation of the results.

THE SOLUTION OF POLYNOMIAL EQUATIONS WITH THE AID OF THE ELECTROLYTIC TANK

By L. Tasny-Tschiassny* and A. G. Doe*

[*Manuscript received April* 10, 1951]

Summary

A practical method is presented of solving polynomial equations with real or complex coefficients with the aid of an electrolytic tank. The roots may be determined with any desired degree of accuracy by means of an iterative procedure.

A method of generating the electrode currents, and a method of ascertaining root points with the aid of a rotatable quadruple probe assembly and passive phase shifting and differential networks are described. A mathematical investigation of the accuracy to be expected is made, and tests have been carried out in a tank, verifying the theoretical conclusions. The methods developed are applicable to the computation of the complex values of rational functions, for instance, impedance functions.

I. Introduction

Lucas's method of solving polynomial equations with the aid of an electrolytic tank has been known for more than 60 years (Lucas 1888; Boothroyd, Cherry, and Makar 1949). By this method all the real and complex roots of a given polynomial equation for Z† with real coefficients L_s

$$G(Z)=\sum_{s=0}^{s=n} L_s Z^s=0, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

can be determined. The accuracy is very limited, and this is apparently the reason why, so far, nothing has become known about practical utilization of Lucas's ideas.

Recently Bloch (1948) has brought forward a suggestion for extending Lucas's method to equations with complex coefficients, but he has not published any indication of the way in which his idea could be used in practical applications.

In the present paper the fact is utilized that Bloch's extension permits a development of Lucas's work into a valuable working method for the solution of polynomial equations with real or complex coefficients, because it opens the road to an iterative procedure, and consequently to unlimited accuracy.

The substitution of

$$Z=u.Z'+Z_0, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

in (**1**), where u is a suitable real and Z_0 a suitable complex number, leads to a polynomial equation with complex coefficients for a new unknown, Z'. If the complex quantities Z, Z', and Z_0 are considered lengths, they define points in an electrolytic tank, and by suitable choice of u and Z_0 the scale and origin of a Cartesian system of coordinates referring to the tank may be changed as desired. From the approximate roots obtained by the first application of the methods discussed in this paper $Z_0=x_0+jy_0$ may be chosen near one of the approximate values found for the complex roots, and u chosen so that the root is conveniently located in the tank. Then Z' may be determined with greater absolute accuracy than Z. The process may be repeated several times to determine the root with ever-increasing accuracy.

---

* Department of Electrical Engineering, University of Sydney.

† Capital letters in this paper will, as far as practicable, stand for complex quantities and small letters for real ones.

The main object of the present paper is the presentation of a practicable system of attacking the problems in question. The principles of the system were tested by actual measurements in a tank, and the results were very encouraging.

II. MATHEMATICAL BASIS OF THE MODIFIED LUCAS-BLOCH METHOD

The Lucas-Bloch method modified by the authors for the purpose of obtaining a greater flexibility exhibits two characteristic features: (i) an auxiliary rational function is used, the zeros of which correspond to the zeros of $G(Z)$; and (ii) this auxiliary function is represented by an electrical analogue in the electrolytic tank.

(a) *Auxiliary Rational Function*

One selects m complex numbers

$$Z_s=x_s+jy_s\,; \quad s=1 \text{ to } s=m>n,$$

which, for the purpose of the discussion, are not roots of the equation.* To each Z_s a complex number Z'_s corresponds according to equation (**2**).

One defines

$$F(Z')=\prod_{s=1}^{s=m}(Z'-Z'_s)=u^{-m}\cdot\prod_{s=1}^{s=m}(Z-Z_s), \quad \ldots\ldots\ldots\ldots \quad (3)$$

and

$$H(Z')=\frac{G(Z)}{F(Z')}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

$H(Z')$ can be expressed as the sum of partial fractions

$$H(Z')=\sum_{s=1}^{s=m}\frac{A_s}{Z'-Z'_s}, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

where

$$A_s=a_s^{(r)}+ja_s^{(i)}=\frac{G(Z_s)}{B_s(Z'_s)}, \quad \ldots\ldots\ldots\ldots\ldots \quad (6)$$

and $B_s(Z'_s)$ is given by the definition of the function

$$B_s(Z')=\prod_{q=1 \text{ to } (s-1)}^{q=(s+1) \text{ to } m}(Z'-Z'_q). \quad \ldots\ldots\ldots\ldots \quad (7)$$

The symbol $\prod_{q=1 \text{ to } (s-1)}^{q=(s+1) \text{ to } m}$ denotes that q assumes all integral values between 1 and m with the exception of s. If y_0 is constant, i.e. the points Z'_s lie on a straight line parallel to the real axis, the quantities $B_s(Z'_s)$ become real.

---

* As shown in Appendix I this restriction is unnecessary, because the procedure remains unchanged if some of the quantities Z_s happen to be roots of equation (**1**).

By eliminating in (1) to (5) Z', $G(Z)$, $F(Z')$, and $H(Z')$, one obtains two power series in Z. The comparison of the first coefficients shows that

$$\left.\begin{aligned}&\Sigma A_s=0, \text{ if } m\geqslant n+2,\\ &\Sigma A_s=u^{n-1}\cdot L_n, \text{ if } m=n+1.\end{aligned}\right\} \quad \ldots\ldots\ldots\ldots \quad (8)$$

(b) The Electrolytic Tank Analogue

Let sinusoidal currents, the r.m.s. values of which are represented by the complex numbers

$$I_s=i_s^{(r)}+j.i_s^{(i)}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (9)$$

be driven into m thin cylindrical electrodes placed at the points Z'_s of the tank with their axes vertical. The instantaneous values $i_s(t)$ of these currents are equal to $\sqrt{2}$ times the projection on the real axis of the vectors representing these complex numbers, viz.

$$i_s(t)=i_s^{(r)}\,.\,\sqrt{2}\,.\,\cos(\omega t)-i_s^{(i)}\,.\,\sqrt{2}\,.\,\sin(\omega t). \quad \ldots\ldots \quad (10)$$

Let

$$C=c^{(r)}+j\,.\,c^{(i)}=c\,.\,\varepsilon^{j\gamma} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (11)$$

be a suitable constant, and let the currents I_s comply with the proportionality condition

$$I_s=C\,.\,A_s. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (12)$$

If $m\geqslant n+2$, the system of currents is self-contained (see equation (8)), if $m=n+1$ a return electrode at theoretically the point $Z'=\infty$ must be provided for.

The space vector of the current density at the point Z' that is caused by a current $i_s(t)$ at the point Z'_s can be represented by a complex number $D(Z', t)$. If h is the height of the electrolyte, and the symbol * denotes the conjugate complex number, one obtains from fundamental considerations†

$$D(Z', t)=\frac{i_s(t)}{2\pi h\,.\,(Z'-Z'_s)^*}. \quad \ldots\ldots\ldots\ldots\ldots \quad (13)$$

The instantaneous values of the space vectors of the current densities $D_r(Z',t)$ and $D_i(Z',t)$ which, according to equations (11) to (13), correspond to the separate existence of all real parts $a_s^{(r)}$ and imaginary parts $a_s^{(i)}$ respectively of the complex numbers A_s, are given by

$$\left.\begin{aligned}D_r(Z',t)&=\frac{c\sqrt{2}\,\cos\,(\omega t+\gamma)}{2\pi h}\sum_{s=1}^{s=m}\frac{a_s^{(r)}}{(Z'-Z'_s)^*},\\ D_i(Z',t)&=-\frac{c\sqrt{2}\,\sin\,(\omega t+\gamma)}{2\pi h}\sum_{s=1}^{s=m}\frac{a_s^{(i)}}{(Z'-Z'_s)^*}.\end{aligned}\right\} \quad \ldots\ldots\ldots \quad (14)$$

---

† The correctness of equation (13) is immediately evident, if one checks separately the magnitude and the phase angle of $D(Z', t)$. The magnitude of the current density $|D(Z', t)|$ multiplied by the area of the cylindrical surface $2\pi h\,|(Z'-Z'_s)|$ equals the current $i_s(t)$, and the direction of the current density is radially outwards from the conductor.

Let $\overline{D_r(Z')}$ and $\overline{D_i(Z')}$ be the r.m.s. values† of the space vectors $D_r(Z',t)$ and $D_i(Z',t)$ respectively.

$$\left.\begin{aligned}\overline{D_r(Z')}&=\frac{c}{2\pi h}\cdot\sum_{s=1}^{s=m}\frac{a_s^{(r)}}{(Z'-Z'_s)^*},\\ \overline{D_i(Z')}&=\frac{c}{2\pi h}\cdot\sum_{s=1}^{s=m}\frac{a_s^{(i)}}{(Z'-Z'_s)^*}.\end{aligned}\right\}\quad\ldots\ldots\ldots\ldots (15)$$

From equations (5) to (15) follows:

$$H(Z')=\frac{2\pi h}{c}\cdot[\overline{D_r^*(Z')}+j\,.\,\overline{D_i^*(Z')}].\quad\ldots\ldots\ldots\ldots (16)$$

For roots of (1) it is evident from (4) that $H(Z')=0$, and one obtains

$$\overline{D_i(Z')}=-j\overline{D_r(Z')}.\quad\ldots\ldots\ldots\ldots\ldots\ldots (17)$$

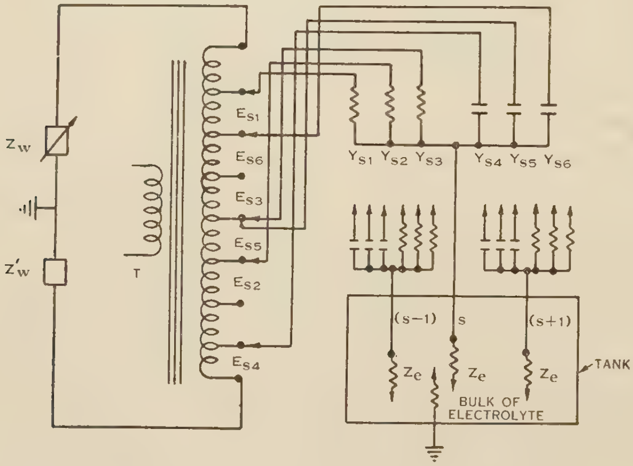


Fig. 1.—Generation of electrode currents.

The actual current density at the individual points of the tank corresponds to a superposition of $D_r(Z',t)$ and $D_i(Z',t)$. An inspection of equations (14), (15), and (17) and the additional mathematical treatment in Subsection IV (*b*) show that the actual current density in the tank at points corresponding to roots of (1), and at these points only, is circularly polarized.

III. The Generation of the Currents Impressed on the Electrodes

(*a*) *The Principle of the Method*

A convenient method of generating the electrode currents apparently not previously used is illustrated in Figure 1. *T* is a transformer of negligible internal impedance having $(2p-1)$ secondary tappings to give voltages E/p,

---

†The bar on top of a quantity denotes in this paper the r.m.s. value. This deviates from the usual practice. For instance, $\overline{D_i(Z')}$ is the r.m.s. value of the vector $D_i(Z',t)$ which has a constant direction, but varies in magnitude with time. $d_i(Z')$ is the length of the vector $\overline{D_i(Z')}$.

$2E/p$, $3E/p$, ... E, $-E/p$, $-2E/p$, $-3E/p$, ... $-E$ with respect to the centre tap. Let each electrode s be provided with a set of k external admittances Y_{sq} connected at one end to the electrode. For fixed q and varying s the Y_{sq} s are given the same nominal values Y_q. The free end of each of the admittances is connected to a suitable tapping of the transformer T, giving a nominal voltage E_{sq} with respect to the centre tap. If $m=n+1$, let one additional electrode at theoretically $Z'=\infty$ be included as an electrode s, so that the sum of all electrode currents is always zero.

By appropriate selection of values E_{sq} a current value I_{s0} nearly proportional to the value desired to leave an electrode can be adjusted according to the formula

$$I_{s0}=\sum_{q=1}^{q=k}(E_{sq}Y_q). \qquad \text{(18)}$$

If, for instance, $p=5$, and six admittances are provided for, of the values $Y_1=1000g$, $Y_2=100g$, $Y_3=10g$, $Y_4=1000jg$, $Y_5=100jg$, $Y_6=10jg$, it is possible to adjust I_{s0} in steps of 2 units, if 1110 units are the maximum adjustable value of I_{s0}. One arranges that

$$\sum_{s=1}^{s=m'} I_{s0}=0, \qquad \text{(19)}$$

where $m'=m+1$ or m, according to whether the point ∞ is included or not.

For small diameters and large mutual distances of the electrodes the seat of the main part of the mutual impedances between them is in the vicinity of the electrodes. Hence the tank can, in good approximation, be represented by a star of equal impedances where $\mathbf{Z}_e$ is the impedance between one electrode and the bulk of the electrolyte.

Let I_s be the current impressed on the electrode s, and E_0 the potential of the centre tap of the transformer with respect to the bulk of the electrolyte, if all admittances and transformer voltages are exactly at their correct values. Thus one obtains

$$I_s=\sum_{q=1}^{q=k}[(E_{sq}+E_0-I_s\mathbf{Z}_e)\,.\,Y_q]. \qquad \text{(20)}$$

Solving for I_s gives

$$I_s=\frac{1}{K}\cdot\sum_{q=1}^{q=k}(E_{sq}Y_q)+\frac{1}{K}\cdot E_0\cdot\sum_{q=1}^{q=k}Y_q, \qquad \text{(21)}$$

where

$$K=1+\mathbf{Z}_e\cdot\sum_{q=1}^{q=k}Y_q. \qquad \text{(22)}$$

If the magnitude of $\mathbf{Z}_e$ is small compared with that of $\sum_{q=1}^{q=k} Y_q$, slight variations in $\mathbf{Z}_e$ for the individual electrodes introduce only second order errors. This is desirable. Hence, unless the electrodes are symmetrically arranged, the resistivity of the electrolyte should be sufficiently small.

Computing $\sum_{s=1}^{s=m'} I_s$ from (**21**), equating it to zero, and considering equations (**18**) and (**19**) show that $E_0=0$ and the second term on the right-hand side of (**21**) vanishes. Consequently,

$$I_s=\frac{1}{K}\cdot I_{s0}. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad \mathbf{(23)}$$

(*b*) *Practical Realization*

When the admittance arrangement shown in Figure 1 was set up, considerable difficulties arose owing to the existence of small high order harmonics in the supply voltage and the ensuing appreciable high order harmonics content of the condenser currents. These harmonics can be rendered negligible by the expedient illustrated in Figure 2. Resistances of nominal values equal to the nominal values of the reactances of the capacitances at supply frequency were inserted in series with the capacitances. This reduces the capacitive current to half the value of the capacitive current without the series resistor, and a resistive current exists which must be compensated for by the adjustment of the resistive currents in the parallel resistance arms 1, 2, and 3 (Fig. 2).

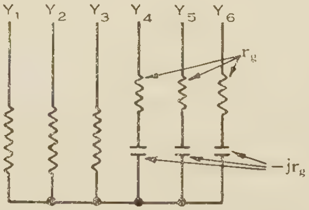


Fig. 2.—Modification to minimize harmonics.

IV. The Detection of Points with Circularly-polarized Field

(*a*) *Inherent Difficulties of the Detection*

The detection of the points where a circularly-polarized field gradient exists, by the use of a twin probe assembly rotatable around a vertical axis, appears very promising at first glance. One connects a vibration galvanometer across the probes and explores the tank for points at which the indication of the detector is independent of the angular position of the probe assembly. Unfortunately, without some basic modifications the procedure described is impracticable for three reasons :

(i) If the intensity of the revolving field gradient is considerable, the accuracy in detecting its fluctuations with the direction is insufficient to determine the position of the root closely enough.

(ii) The sensitivity of a vibration galvanometer does not suffice without the use of an amplifier. Every amplifier has one point of its circuit connected to earth either directly or through an impedance. Consequently, owing to leakages and stray capacities to earth of the parts of the transformer winding, some current enters the tank through the probe assembly and disturbs the measurements, because $\sum_{s=1}^{s=m'} I_s$ is no longer equal to zero, and the results of Section III are no longer valid.

(iii) The voltages to be detected in the tank are so small that the stray voltages induced in the leads and in the amplifier can be of a greater order of magnitude than these voltages, even if careful screening is applied.

(*b*) *Methods of Overcoming the Difficulties*

The first difficulty mentioned in Subsection IV (*a*) can be overcome by the use of a quadruple probe assembly, enabling the simultaneous pick-up of the voltage gradients, and thus the current densities, in two perpendicular directions and, as will be explained presently, by the insertion of a phase shifting and a differential network. The phase shifting and differential networks are connected in cascade (see the block diagram, Fig. 3), and the output of the differential network is amplified by a battery-operated amplifier and detected by a vibration galvanometer. The phase shifting and differential networks can be made up of passive components, i.e. without the use of valves, and no requirements are necessary as to the constancy of the gain of the amplifier.

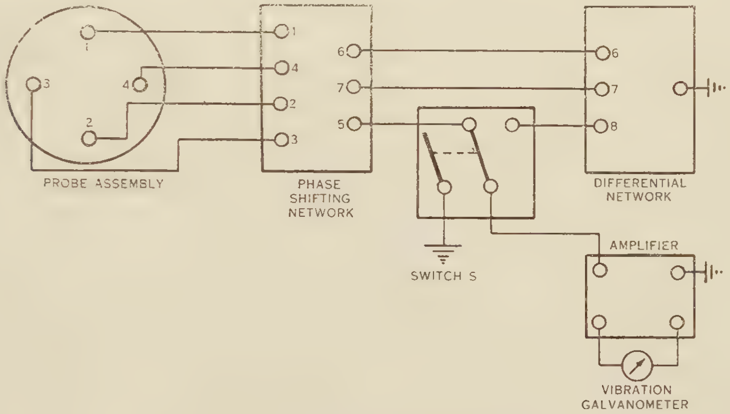


Fig. 3.—Block diagram of apparatus.

In Figure 4 the vectors $\overline{D_r(Z')}$ and $\overline{D_i(Z')}$ represent the complex current densities given by equations (**15**), and $\overline{d_r(Z')}$ and $\overline{d_i(Z')}$ their magnitudes. Let φ be the angle between these vectors. One obtains from equations (**14**) and trigonometrical considerations that the instantaneous value $d_{12}(Z',t)$ of the current density in any direction given by the angle θ (see Fig. 4), and resulting from the simultaneous action of $D_r(Z',t)$ and $D_i(Z',t)$, fluctuates according to

$$\begin{aligned} d_{12}(Z',t) &= \overline{d_r(Z')} \cdot \sqrt{2} \cdot \cos(\omega t+\gamma) \cdot \cos\theta \\ &\quad -\overline{d_i(Z')} \cdot \sqrt{2} \cdot \sin(\omega t+\gamma) \cdot \cos(\theta-\varphi). \quad \ldots . \ (\mathbf{24}) \end{aligned}$$

Let the instantaneous value of the current density in a direction perpendicular to that given by the angle θ, i.e. in the direction given by the angle $(90°+\theta)$, be $d_{34}(Z',t)$. One obtains

$$d_{34}(Z',t)=-\overline{d_r(Z')}\,.\,\sqrt{2}\,.\,\cos(\omega t+\gamma)\,.\,\sin\theta$$
$$+d_i(Z')\,.\,\sqrt{2}\,.\,\sin(\omega t+\gamma)\,.\,\sin(\theta-\varphi).\quad\ldots\ (25)$$

If the phase shifting network makes the voltage that corresponds to $d_{12}(Z',t)$ lag by $\alpha°$ and the voltage that corresponds to $d_{34}(Z',t)$ lag by $(\alpha+90°)$ and the differential network computes at any instant the difference of the two shifted voltages, then the output of the differential network is proportional to a fictitious current density $d_0(Z',t)$ given by

$$d_0(Z',t)=\overline{d_r(Z')}\,.\,\sqrt{2}\,.\,\cos(\omega t+\gamma+\alpha+\theta)$$
$$+d_i(Z')\,.\,\sqrt{2}\,.\,\cos(\omega t+\gamma+\alpha+\theta-\varphi+90°).\quad\ldots.\ (26)$$

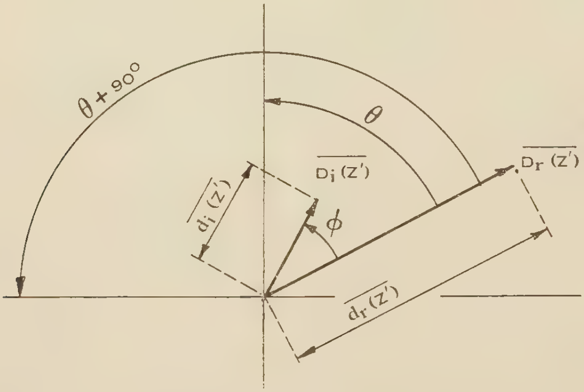


Fig. 4.—Diagram of current density space vectors.

A closer investigation of equation (26) reveals that the r.m.s. value $\overline{d_0(Z')}$ of the fictitious current density $d_0(Z',t)$ is proportional to the magnitude of the vector $[D_r^*(Z')+jD_i^*(Z')]$ where the * stands for the conjugate complex values. Consequently, one obtains from equation (16) that

$$|\mathrm{H}(Z')|=\frac{2\pi h}{c}\cdot\overline{d_0(Z')}.\quad\ldots\ldots\ldots\ldots\ldots\ (27)$$

Hence, the output of the differential network is a measure that shows how far one is away from a root. The phase of the output of the differential network varies with the orientation of the quadruple probe assembly, but its magnitude does not.

The second difficulty mentioned above can be overcome by the use of a parallel twin-T network as differential network and the use of a Wagner earth. This will be explained in Subsection IV (*d*).

The third difficulty can be overcome by making the quadruple probe assembly rotatable. Since the phase of the stray pick-up is practically independent of the orientation of the probe assembly, but the output voltage of the differential network changes its phase, a satisfactory measuring procedure is to rotate the probe assembly continuously and to explore the tank for points at which the indication of the detector is independent of the angular position of the probe assembly and becomes a minimum at the same time. These points correspond to roots of the equation. An additional procedure to eliminate the influence of stray pick-up is given in Subsection IV (*g*).

(*c*) *The Phase Shifting Network*

The phase shifting network consists of four equal resistors r_1 and four equal condensers c_1 connected to the four probes 1 to 4 of the quadruple probe assembly in the way shown in Figure 5. A star consisting of four equal resistors r_2 is added to obtain a point 5, the potential of which is identical with the potential of the centre of the probe assembly. For the sake of convenience this star will be considered a part of the phase shifting network. The differential network will be connected to the points 5, 6, and 7, and it will be designed in such a way that its output will be proportional to the difference in potential of the points 6 and 7. The branches 2–3 and 4–1 are inserted purely in order to obtain complete symmetry.

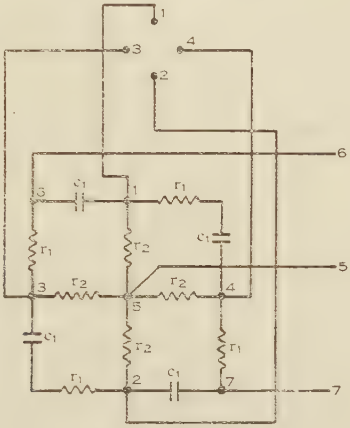


Fig. 5.—Phase shifting network.

For the mathematical treatment of the conditions the assumption is necessary that the presence of a probe of finite diameter does not disturb the field set up by the influence of the currents emanating from the electrodes and the other probes. No serious error is expected to result from this assumption.

Let V_1', V_2', V_3', V_4' be the potentials of the four probes 1, 2, 3, 4 respectively with respect to earth, if the phase shifting network is disconnected, and V_1, V_2, V_3, V_4 these potentials, if it is connected, but no further connection is made. If the probe assembly is not too close to any electrode, and the resistivity of the electrolyte is sufficiently low, the arrangement can be considered as being symmetrical. Consequently

$$\frac{V_1-V_2}{V_1'-V_2'}=\frac{V_3-V_4}{V_3'-V_4'}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (28)$$

Let V_6' and V_7' be the potentials of the points 6 and 7 with respect to earth, if the differential network is not connected.

A simple calculation shows that

$$V_6'-V_7'=\frac{1}{r_1-j/\omega c_1}\left[r_1(V_1-V_2)-\frac{j}{\omega c_1}(V_3-V_4)\right]. \quad \ldots (29)$$

If

$$r_1=\frac{1}{\omega c_1}, \quad \ldots\ldots (30)$$

the assumptions made in Subsection IV (*b*) on the action of the phase shifting and differential networks are complied with by the right-hand side of equation (29), and therefore the magnitude of $V_6'-V_7'$ is proportional to the fictitious current density $\overline{d_0(Z')}$. If ρ denotes the resistivity of the electrolyte and e' the distance of the probes 1 and 2, and 3 and 4 respectively, expressed in the same length units in which Z' is measured, one obtains

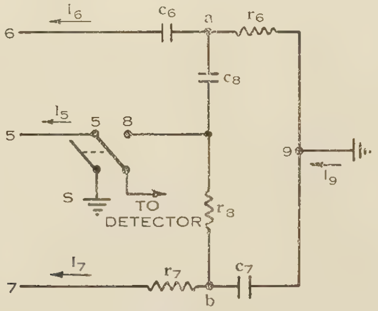


Fig. 6.—Differential network.

$$|V_6'-V_7'|=\frac{\rho e' \cdot \overline{d_0(Z')}}{\sqrt{2}}. \quad \ldots\ldots (31)$$

Hence, $V_6'-V_7'=0$ is the condition for the centre of the probe assembly being positioned at a root of equation (1).

(*d*) *The Differential Network*

The differential network (Fig. 6) consists of a parallel twin-T formed by the condensers c_6, c_7, c_8, and the resistors r_6, r_7, r_8. It is designed in a symmetrical way so that

$$\left.\begin{aligned} r_6&=\frac{1}{\omega c_7},\\ r_7&=\frac{1}{\omega c_6},\\ r_8&=\frac{1}{\omega c_8},\end{aligned}\right\}. \quad \ldots\ldots (32)$$

Let V_6 and V_7 be the potentials of the points 6 and 7 of the phase shifting network, if the differential network is connected to it. If $V_6'-V_7'=0$, then $V_6-V_7=0$.

If $V_6 - V_7' \neq 0$, and consequently, according to Thevenin's theorem $V_6 - V_7 \neq 0$, the output of the differential network, i.e. the potential of the point 8 with respect to earth, is different from zero, as its analysis (see Appendix II) shows, provided that in the design the condition

$$\frac{1}{r_6} = \frac{1}{r_7} + \frac{1}{r_8} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (33)$$

is complied with. The result is the equation

$$V_8 = \tfrac{1}{2} j \frac{r_6}{r_7} (V_6 - V_7), \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (34)$$

which expresses the above statement.

(e) *The Wagner Earth*

The Wagner earth consists of the two impedances $\mathbf{Z}_w$ and $\mathbf{Z}'_w$ (Fig. 1) and a double pole double throw switch S (Figs. 3 and 6). The impedances $\mathbf{Z}_w$ and $\mathbf{Z}'_w$ of the Wagner earth auxiliary bridge arms are adjusted to zero detector indication with the switch S in its left-hand position. For the exploring measurements the switch S is put in its right-hand position and the point 5 is earthed directly, but, owing to the previous Wagner earth adjustment, the earth current remains zero.

A maladjustment of the Wagner earth auxiliary bridge does not influence to any great extent the working of the detecting arrangement : it causes small currents I_5 and I_9 to enter the probe assembly from the points 5 and 9 (see Fig. 6). The current I_5 divides equally between the probes 1 to 4 and influences the voltages V_1 to V_4. It can easily be seen that the potential differences $(V_1 - V_2)$ and $(V_3 - V_4)$ are not affected by the presence of the current I_5. If the magnitude of the impedance of the shunt arm formed by the series combination of r_8 and c_8 is much greater than r_6 (and this can easily be arranged for), the two currents I_6 and I_7 (see Fig. 6) will be practically equal. Consequently, the currents established in the condensers c_1 in the branches 6–1 and 7–2, as well as in the resistors r_1 in the branches 6–3 and 7–4, will be equal respectively, so that the influence of the current I_9 on the voltages V_1 to V_4 will be such that the differences $(V_1 - V_2)$ and $(V_3 - V_4)$ will not be affected. Hence, the only practical consequence of a maladjustment of the Wagner earth auxiliary bridge arms is the fact that the sum of the currents I_s given by equation (**21**) is no longer equal to zero, and equation (23) does not hold any longer. But this influence is small. Experience in the use of the tank confirmed the expected insensitivity of the arrangement to maladjustment of the Wagner earth. It needs only occasional adjustment while the probe assembly is moved across the tank, and this simplifies the work considerably.

(f) *Adjustment of the Networks and Modifications Required*

The detecting network is adjusted in two steps. First, voltage of supply frequency is applied between the connected terminals 6 and 7 on the one hand and terminal 9 on the other hand of the differential network while the other parts of the detecting network are disconnected. By changing slightly two appropriate components, for instance r_6 and c_7, zero output between 8 and 9 can be achieved.

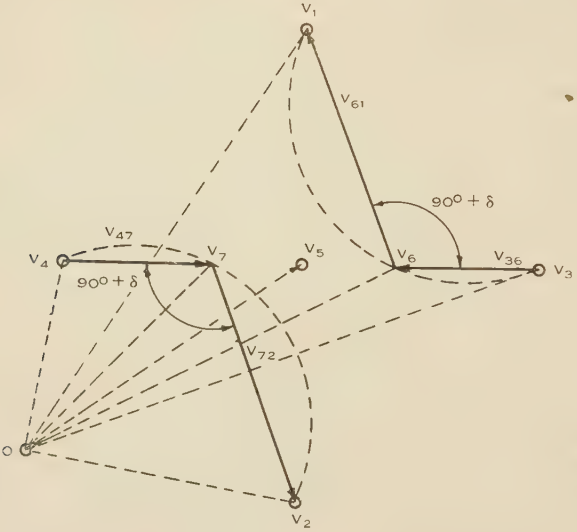

V1
V61
90° + δ
V36
V3
V6
V5
V47
V4
V7
90° + δ
V72
O
V2

Fig. 7.—Potential diagram of the phase shifting network.

Then the total detecting device is adjusted to zero or constant near-zero output between 8 and 9, while the probe assembly is immersed in a tank at a point of circularly-polarized field and rotated. If the condensers c_1 have losses, this adjustment is not possible without further provisions. Figure 7 shows the potential diagram of the phase shifting network, i.e. the diagram of the end-points V_1, V_2, V_3, V_4, V_5, V_6, and V_7 of the voltage vectors from a suitable reference

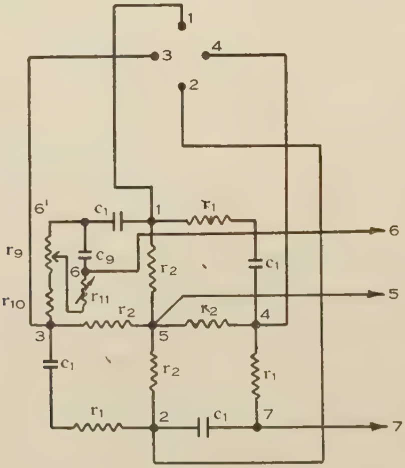

1
3
4
2
6'
c1
1
r1
6
r9
c9
r2
c1
6
r10
r11
r2
r2
4
5
3
5
c1
r2
r1
r1
2
c1
7
7

Fig. 8 (*a*).—Modified phase shifting network.

point 0, drawn as dotted lines. The difference vectors V_{36} and V_{61}, as well as the difference vectors V_{47} and V_{72}, include an angle $(90+\delta)°$ respectively, where δ is the dielectric loss angle of the condensers c_1. The points V_6 and V_7 do not coincide. Various means are available for achieving coincidence. One is to arrange the probe assembly not exactly in a square, but to make the distance between probes 3 and 4 slightly shorter than that between probes 1 and 2. Another would be to make the angle between the lines connecting probes 1 and 2, and 3 and 4 respectively slightly different from 90°. Then one resistor (for instance r_1 in the branch 6–3) and a small resistor connected in series with the corresponding condenser c_1 in the branch 1–6 are to be adjusted. The authors left the probes arranged in a square and used a third method, viz. the network shown in Figure 8 (a), in which the components r_9 and r_{10} were adjusted. Figure 8 (b) shows the potential diagram idealized under the assumption that the impedances of r_{11} and c_9 are very high. For a fixed position of r_9 the locus of V_6, as r_{11} is varied, lies on the semi-circle shown in Figure 8 (b). The diameter of the circle can be varied by adjustment of r_9 so that V_6 may be adjusted to coincide with V_7.

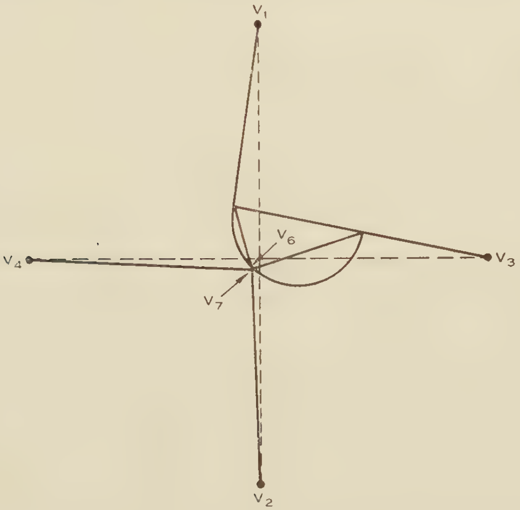


Fig. 8 (b).—Vector diagram for modified phase shifting network.

To produce a circularly-polarized field a separate round tank of smaller dimensions was used, in which four wall electrodes arranged in a square were energized by currents proportional to $+1$ and -1 in one diagonal and $+j$ and $-j$ in the other diagonal. The field in the centre of the tank is circularly polarized. Since the arrangement corresponds to the equation $G(Z)=(Z-j)^2$, the root in the centre is double. As can be seen from Subsection VI (d), slight errors in the position of the probe assembly do not appreciably change the character of the field, so that the arrangement with a double root is particularly well suited for adjusting purposes.

All the procedures described are such that deviations in the arrangement of the probes from their intended arrangement are automatically compensated by the adjustment.

(g) *Elimination of the Influence of Residual Stray Fields*

The authors used in their measurements leads with a copper wire braiding as shield, and the networks and amplifier were screened by cases made up of mild steel of about 0·040 in. thickness. The amplifier was battery operated. No further precautions were taken. Caused by stray pick-up, the vibration galvanometer exhibited at root points a residual deflection (not altering when the probe was rotated), which exceeded considerably the variations of this deflection caused by the current density in the tank when the probe assembly was rotated at a point of the tank, say 1 in. away from the root point. Measurements made under these conditions would not have been accurate enough. The residual deflection caused by the stray fields could quite easily be reduced to a magnitude not affecting the accuracy of the measurements at all by injecting in series with the vibration galvanometer a small voltage of appropriate magnitude and phase. The authors used for this purpose the balancing voltage of an available portable Gall coordinate type A.C. potentiometer; but any other auxiliary network producing a voltage floating with respect to earth, and of adjustable magnitude and phase, would have been suitable.

V. TANK OF FINITE DIMENSIONS

(a) *Possible Methods*

The use of a tank of finite dimensions introduces serious errors, unless special arrangements are made. As was pointed out by Boothroyd, Cherry, and Makar (1949), the equivalent of a tank of infinite dimensions can be achieved: (i) by the use of a double-layer tank*; and (ii) by conformal transformation as indicated in Subsection V (*b*).

(b) *Conformal Transformation of the Upper Half-Plane into the Interior of a Circle*

The basis of the experimental work in the electrolytic tank can be made rigorous by arranging all points Z'_s on the line $y'_s=0$, and by transforming the upper half-plane of the variable Z' into the interior of a circle (see Fig. 9). This circle is represented by a circular tank of diameter d with a wall of insulating material to which all electrodes are attached. The transforming function is

$$Z''=x''+jy''=-\frac{d}{Z'+j}, \quad \dots\dots\dots\dots\dots\dots \quad (35)$$

where $Z''=x''+jy''$ has the meaning shown in Figure 10.

---

* For information on the double-layer tank the reader is referred to the paper by Boothroyd, Cherry, and Makar.

The minimum number of currents to be impressed is $(n+2)$. Should $m=n+1$, the $(n+2)_{nd}$ current is at the point $Z''_{\infty}=0$, corresponding to $Z'=\infty$ and has the value determined by the negative sum of all other $(n+1)$ electrode currents.

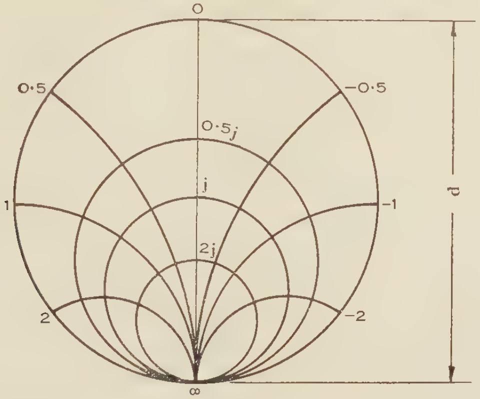


Fig. 9.—Conformal transformation for a circular tank.

(c) *Iterative Procedure in the Circular Tank*

After an approximate value (x_a+jy_a) of a root has been found, an upper limit of the possible error is either known from previous experience or can be estimated (see Section VI). One makes the real part x_0 of Z_0 approximately equal to x_a and selects the value y_0 in such a way that the possible values of the root lie in the interior of the tank.* Then u is given a value such that the most probable value of the root, i.e. in the first instance the value (x_a+jy_a), is near the centre of the tank. This is desirable to achieve the symmetry assumed in Section III, but there is another reason. By a simple calculation it can be

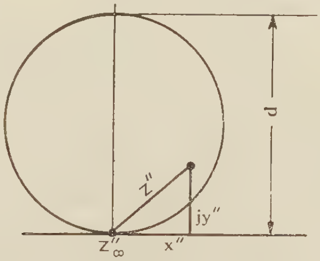


Fig. 10.—Coordinates in a circular tank.

---

* It is worth while mentioning that an exchange of the domains represented by the interior and exterior of the tank, i.e. the representation of the lower half-plane of Z' by the interior of the tank, can be easily brought about by changing the sign of all resistive components of the electrode currents. If by any chance the value y_0 required for the transformation was not appropriately selected, this artifice may save time.

shown that for $x=x_a$ the value $\left|\dfrac{dZ''}{dZ}\right|$, if considered a function of u, has a maximum at $Z''=j^{d/2}$, i.e. the centre of the tank. Hence, for variable u and a given small change of Z the position of the probe assembly changes most if arranged near the centre of the tank.

The example in Section VIII illustrates the use of the circular tank.

VI. Estimate of Errors

(a) Sources of Errors

The errors that occur in the determination of a root have mechanical causes (inaccurate position of the electrodes, inclination of the bottom plate of the tank to the horizontal) and electrical causes (the electrode currents are not accurately set, the frequency of the supply deviates from its nominal value, contact resistances are present). Further errors are caused by inaccuracies in determining the position of the probe assembly and by the limited sensitivity of the detecting device.

If copper sulphate solution (Hansen and Lundstrom 1945) of comparatively high resistivity is used and the electrodes are of copper, and clean, the influence of the contact resistances can be neglected. A deviation of the supply frequency changes the setting of the electrode currents and upsets slightly the adjustment of the detecting network, but the latter effect has no practical consequences. An inclination of the bottom plate of the tank is to a great extent compensated by the fact that the relative values of the partial current densities at a certain point of the tank are little influenced by deviations of the height of the electrolyte at the electrodes, as long as the electrode currents remain constant.

(b) Consequence of Inaccurate Transformer Voltages and Inaccurate Values of the Admittances Y_q

It was assumed in Section III that the transformer voltages E_{sq} and the admittances Y_q do not deviate from their nominal values. Let $I_s+\triangle I_s$, $E_0+\triangle E_0=\triangle E_0$, $E_{sq}+\triangle E_{sq}$, and $Y_q+\triangle Y_{sq}$ be the actual values of the quantities I_s, $E_0=0$, E_{sq} and Y_q. For an estimate of the error the quantity K, which does not differ greatly from unity, can be considered as remaining constant. By taking the total differential of equation (21) one obtains

$$\triangle I_s=\frac{1}{K}\cdot\sum_{q=1}^{q=k}(\triangle E_{sq}\cdot Y_q+E_{sq}:\triangle Y_{sq})+\frac{1}{K}\triangle E_0\cdot\sum_{q=1}^{q=k}Y_q. \quad \ldots\ldots (36)$$

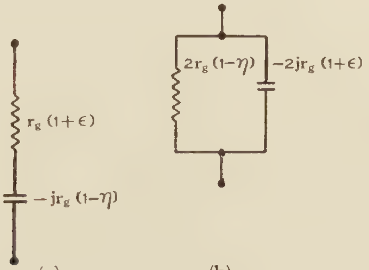
Computing $\sum\limits_{s=1}^{s=m}\triangle I_s$ from (36) and equating it to zero leads to an equation for $\triangle E_0$. If the resulting value of $\triangle E_0$ is substituted in (36) one obtains

$$\triangle I_s=\frac{1}{K}\cdot\left[\frac{m'-1}{m'}\cdot\triangle I_{s0}-\frac{1}{m'}\cdot\sum_{q=1\text{ to }(s-1)}^{q=(s+1)\text{ to }m'}\triangle I_{q0}\right], \quad \ldots\ldots\ldots\ldots (37)$$

where

$$\triangle I_{q0}=\sum_{r=1}^{r=k}(\triangle E_{qr}\cdot Y_r+E_{qr}\cdot\triangle Y_{qr}). \quad \ldots\ldots\ldots\ldots\ldots (38)$$

The maximum value of $\triangle E_{qr}$ is known from the design data of the transformer used. As to the estimate of the error $\triangle Y_{qr}$, let us assume that the resistance r_g of Figure 2 has a small relative error ε and the corresponding reactance of the series condenser a small relative error η. Figure 11 (*a*) shows the actual circuit. A simple calculation leads to the component values entered in Figure 11 (*b*) for the equivalent parallel circuit. An error in the value of the actual series resistance is reflected as an error in the equivalent shunt capacity and vice versa.



(a) (b)

Fig. 11.—Errors in equivalent series and parallel combination of resistance and capacitance.

(*c*) *Consequence of Inaccurate Positions of the Electrodes, Inaccurate Setting of the Currents, and Limited Sensitivity of the Detector*

Let $(Z_s+\triangle Z_s)$ correspond to the position of the electrodes, $(I_s+\triangle I_s)$ be the values of the electrode currents, $(Z_r+\triangle Z_r)$ the value of the apparent root ascertained by the measurements, and $\triangle H$ the deviation of $H(Z')$ from zero caused by the limited sensitivity of the detector, where Z_s, I_s, Z_r, and zero respectively are the correct values. It is further assumed that the electrode placed at the point $Z'=\infty$ is in its correct position and the quantities $\triangle Z_s$ are to be ascertained from this viewpoint. Then a possible error in the position of the electrode at $Z'=\infty$ can be considered an error in the determination of the position of the probe assembly.

Considering $\triangle H$ a total differential one obtains

$$\triangle H=\frac{\partial H}{\partial Z}\cdot\triangle Z_r+\sum_{s=1}^{s=m}\frac{\partial H}{\partial I_s}\cdot\triangle I_s+\sum_{s=1}^{s=m}\frac{\partial H}{\partial Z'_s}\cdot\triangle Z'_s, \quad \ldots\ldots (39)$$

where $\partial H/\partial Z$, $\partial H/\partial I_s$, and $\partial H/\partial Z'_s$ refer to $Z=Z_r$.

From (2), (5), and (12) one obtains

$$\frac{\partial H}{\partial I_s}=\frac{1}{C}\cdot\frac{1}{Z'_r-Z'_s}, \quad \ldots\ldots\ldots\ldots\ldots\ldots (40)$$

and

$$\frac{\partial H}{\partial Z'_s}=\frac{A_s}{(Z'_r-Z'_s)^2}. \quad \ldots\ldots\ldots\ldots\ldots\ldots (41)$$

Since $G(Z_r)=0$, one obtains from (4)

$$\frac{\partial H}{\partial Z}=\frac{1}{F(Z'_r)}\cdot\frac{\partial G}{\partial Z'} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (42)$$

where dG/dZ refers to $Z=Z_r$.

An approximate value of $|\triangle H|$ can be found by the use of equations (27), (31), and (34), and the computation of (V_6-V_7) from $(V_6'-V_7')$ by the aid of Thevenin's theorem.

If, in equation (27), $|\triangle H|$ is written for $|H(Z')|$ and $\overline{d_0(Z')}$ is substituted from (31), one obtains

$$|\triangle H|=\frac{2\pi\sqrt{2}}{c}\cdot\frac{h}{\rho e'}\cdot|(V_6'-V_7')|. \quad \ldots\ldots\ldots (43)$$

Using Thevenin's theorem for the voltages $(V_6'-V_7')$ and (V_6-V_7), considering equations (32), (33), and (34), and making the approximation $r_8=\infty$ leads to

$$|\triangle H|=\frac{4\pi\sqrt{2}}{c}\cdot\frac{h}{\rho e'}\cdot\left(1+\frac{r_1}{2r_6}\right)\cdot|V_8|. \quad \ldots\ldots (44)$$

The magnitude of the voltage $|V_8|$ is given by the sensitivity of the detecting device.

If $\triangle Z_{r1}$, $\triangle Z_{r2}$, and $\triangle Z_{r3}$ denote the errors in Z_r caused by the exclusive influence of $\triangle H$, the $\triangle I_s-s$ and the $\triangle Z'_s-s$ respectively, and $\triangle A_s$ are the changes in the quantities A_s that correspond to the changes $\triangle I_s$ according to equation (12), one obtains from a combination of equations (2), (6), (39), (40), (41), (42), and (44)

$$|\triangle Z_{r1}|=\frac{1}{c}\cdot\frac{F(Z'_r)}{dG/dZ}\cdot\frac{4\pi\sqrt{2}\,.\,h}{\rho\,.\,e'}\cdot\left(1+\frac{r_1}{2r_6}\right)\cdot|V_8|, \quad (45)$$

$$\triangle Z_{r2}=-\frac{1}{dG/dZ}\cdot\sum_{s=1}^{s=m}\left[\frac{B_s(Z'_r)}{B_s(Z'_s)}\cdot G(Z_s)\cdot\frac{\triangle A_s}{A_s}\right], \quad \ldots\ldots (46)$$

$$\triangle Z_{r3}=-\frac{1}{dG/dZ}\cdot\sum_{s=1}^{s=m}\left[\frac{B_s(Z'_r)}{B_s(Z'_s)}\cdot G(Z_s)\cdot\frac{\triangle Z'_s}{Z'_r-Z'_s}\right], \quad \ldots (47)$$

where dG/dZ refers to $Z=Z_r$.

Equations (45) to (47) contain the proof that the accuracy of the iterative procedure explained in Section I is unlimited. Provided that the positions of the electrodes and the probe assembly remain the same, the only quantities that do alter in consecutive measurements are C, $G(Z_s)$, $\triangle A_s$, and A_s. Since $\triangle A_s/A_s$ does not change either, the errors $\triangle Z_{r2}$ and $\triangle Z_{r3}$ vary in compliance with the values of $G(Z_s)$. In the neighbourhood of a root the values $G(Z_s)$ tend to become proportional to (Z_s-Z_r), viz. they decrease in consecutive measurements. So do the quantities A_s (equation (6)), and consequently $1/C$ (equation (12)), because the order of magnitude of the values I_s remains the same. Consequently $\triangle Z_{r1}$ decreases in a similar way to $\triangle Z_{r2}$ and $\triangle Z_{r3}$.

(*d*) *Multiple Roots*

For a multiple root the given estimate of errors breaks down because dG/dZ becomes zero. Nevertheless, equations (45) to (47) show that $\triangle Z_{r1}$, $\triangle Z_{r2}$, and $\triangle Z_{r3}$ must be large and that, therefore, Lucas's method is not well suited for locating multiple roots of polynomial equations. This is no serious deficiency in the method, because multiple roots of a polynomial equation can be easily ascertained as roots of the common factor of the given polynomial $G(Z)$ and its derivative dG/dZ.

On the other hand, as mentioned in Subsection IV (f), a multiple root offers certain advantages for adjusting purposes. From equations (**39**) and (**42**) it can be seen that for $dG/dZ=0$ the change of $\triangle H$ with $\triangle Z_r$ is very slow near a multiple root, and consequently, according to equation (**44**), $|V_8|$ will be practically zero within a reasonable domain around a multiple root.

VII. APPARATUS

In this section the data of suitable measuring equipment are given. The tank is of convenient size, the admittances and transformers are procurable at reasonable cost, and the accuracy is generally satisfactory. The experiments referred to later were carried out with similar apparatus made up from existing equipment, but with admittance values and transformer voltages of lesser accuracy.

A circular tank of 19 in. diameter is filled with copper sulphate solution of about 75 ohm-inch resistivity to a depth of 3 in., and circular copper electrodes of $\frac{1}{16}$ in. diameter can be attached to the wall in any suitable position.

The supply transformer supplies voltages of ±20, ±40, ±60, ±80, and ±100 V., with an accuracy of ±0·1 per cent. at a frequency of 50 c/s. ±0·5 per cent. The six admittances used are three resistors 3000 ohms ±0·1 per cent., 30,000 ohms ±1 per cent., and 300,000 ohms ±10 per cent., and three series combinations of condensers and resistances, viz. 1·061 µF. ±0·5 per cent. plus 3000 ohms ±0·5 per cent., 0·106 µF. ±1 per cent. plus 30,000 ohms ±1 per cent., and 0·0106 µF. ±10 per cent. plus 300,000 ohms ±10 per cent. The fictitious series resistances that correspond to the dielectric losses of the condensers are included in the given resistance values. This can be easily accomplished within the given tolerance values with good quality condensers.

The probe assembly consists of four copper wires of 0·040 in. diameter arranged in a square of $\frac{5}{16}$ in. side. The probes are kept in fixed mutual position by several insulating discs, some of which are immersed in the electrolyte. The probe assembly is rotatable in a thrust ball-bearing with vertical axis which can be horizontally moved as required. Flexible leads are attached to the probes. This arrangement makes only a reciprocating rotation of the probe assembly possible, but rotatable contact elements with additional contact resistances are avoided.

The component values of the detecting network are given in the following list :

Component Values of the Detecting Networks

$r_1 = 31{,}830\ \Omega$
$r_2 = 10{,}000\ \Omega$
$r_6 = 35{,}370\ \Omega$
$r_7 = 31{,}830\ \Omega$
$r_8 = 318{,}000\ \Omega$
$r_9 = 25{,}000\ \Omega$
$r_{10} = 25{,}000\ \Omega$
$r_{11} = 100{,}000\ \Omega$

$c_1 = 0{\cdot}1\ \mu\text{F.}$
$c_6 = 0{\cdot}09\ \mu\text{F.}$
$c_7 = 0{\cdot}1\ \mu\text{F.}$
$c_8 = 0{\cdot}01\ \mu\text{F.}$
$c_9 = 0{\cdot}02\ \mu\text{F.}$

One microvolt input voltage at the battery operated two-stage amplifier of 1 megohm input impedance produces about $\frac{1}{4}$ in. deflection on the scale of the vibration galvanometer.

VIII. EXAMPLE

The principles involved are best illustrated on the basis of a simple equation such as

$$G(Z)=Z^3-1=0$$

for which the complex root in the positive half-plane, i.e. the root $Z_r=-0\cdot5+j0\cdot866$ is to be determined. The equipment used corresponds to the one described in Section VII. Let $m'=5$ and $m=4$. The selected points Z'_s determining the positions of four electrodes and the corresponding values of $B_s(Z'_s)$ are given in Table 1. The electrode positions in the circular tank are manifest from Figure 9. The arrangement is purposely not symmetrical, though a symmetrical arrangement would yield better results.

TABLE 1

s	1	2	3	4
Z'_s	-1	$-0\cdot5$	$+0\cdot5$	$+1$
$B_s(Z'_s)$	$-1\cdot5$	$+0\cdot75$	$-0\cdot75$	$+1\cdot3$

For the first trial let $Z_0=0$ and $u=4$. Purposely a value of u is chosen which is rather high, so that the root is not near the centre of the tank, though this position would be preferable. The values required for setting the currents with the exclusive use of the resistors are contained in Table 2. The values of A_s and ΣA_s are multiplied by the factor 17 to make the largest value somewhat less than 1110. Then $I_{s0}/100g$ is made as close as possible to $(17A_s)$. Since

TABLE 2

	s	1	2	3	4	Return electrode
	Z'_s	-1	$-0\cdot5$	$0\cdot5$	1	—
	Z_s	-4	-2	2	4	—
	$G(Z_s)$	-65	-9	7	63	—
	A_s	$43\cdot33$	$-12\cdot00$	$-9\cdot33$	42	$-\Sigma A_s=-64$
	$17A_s$	$736\cdot7$	$-204\cdot0$	$-158\cdot7$	$714\cdot0$	$-1088\cdot0$
Connection of:	$1000g$	80V.	—20V.	—20V.	80V.	—100V.
	$100g$	—60V.	0V.	40V.	—80V.	—80V.
	$10g$	—40V.	—40V.	20V.	—60V.	—80V.
	$\frac{I_{s0}}{100g}$	736	—204	—158	714	—1088

$$g=\frac{1}{3\times10^6} \text{ mhos, } CK=\frac{I_{s0}}{A_s}=17\times100g=5\cdot67\times10^{-4}.$$

Let us assume that under unfavourable conditions a value of $Z_r = -0\cdot6 + j1\cdot00$ was found as a first approximation to the root. As was explained in Subsection V (*c*), the origin and scale are changed according to equation (**2**). x_0 is made equal to $(-0\cdot6)$. If a maximum error of $0\cdot3$ is expected—the calculation carried out later in this Subsection gives a value of $0\cdot21$—one must make $y_0 = 0\cdot7j$ so that $Z_0 = -0\cdot6 + 0\cdot7j$. The condition that the point $Z = -0\cdot6 + 1j$ should correspond to the centre of the tank leads to $u = 0\cdot3$. Table 3 is then drawn up. When computing the values $G(Z_s)$ for complex arguments in the general case, it is advisable to transform equation (**1**) into the form

$$G(Z) = \{([(Z \,.\, L_n + L_{n-1}) \,.\, Z + L_{n-2}] \,.\, Z + L_{n-3})Z + \,.\,.\,.\,.\, + L_1\} \,.\, Z + L_0.$$

Then the root is again located with greater absolute accuracy, as shown below.

To obtain an idea of the magnitude of K, the resistance $2\mathbf{Z}_e$ between opposite electrodes, i.e. those at $Z'_s = -1$ and $Z'_s = 1$, as well as the resistance $2\mathbf{Z}'_e$ between the electrodes closest to each other, i.e. those at $Z'_s = -1$ and $Z'_s = -0\cdot5$, were computed. The resistances $2\mathbf{Z}_e$ and $2\mathbf{Z}'_e$ are double the values found for two parallel cylinders in the Z'-plane whereby the diameters of the cylinders in the Z'-plane are to be computed from the physical diameters of the electrodes in the Z''-plane by the formula

$$\frac{\text{Small length in } Z''\text{-plane}}{\text{Corresponding length in } Z'\text{-plane}} = \left|\frac{dZ''}{dZ'}\right| .$$

One obtains $\mathbf{Z}_e = 51$ ohms and $\mathbf{Z}'_e = 41\cdot8$ ohms, or, according to equation (**17**), K varies less than between $K = 1\cdot0189$ and $K = 1\cdot0155$. The mean value of K is approximately $K = 1\cdot017$. Hence $C = 5\cdot59 \times 10^{-4}$. It is safe to assume that the error introduced by the asymmetry of the arrangement of the electrodes is considerably less than $\pm0\cdot17$ per cent. Consequently, the total error in the adjustment of each current caused by the errors in the transformer voltages, resistance values, and the asymmetry of the electrode arrangement will be less than $\pm0\cdot4$ per cent. of the individual current values. To this figure $\pm0\cdot1$ per cent. of the maximum current value is to be added to account for the impossibility of making $I_{s0}/100g$ exactly equal to $17A_s$.

To obtain an idea of the accuracy conditions let $|V_8| = 1\ \mu$V., and let equations (**45**) to (**47**) be used under the assumption that the position of the root Z_r is accurately known. In practice an approximate value would be used. For the first measurements $(Z_0 = 0,\ u = 4)$ one obtains $Z'_r = -0\cdot125 + j0\cdot2165$. The calculations are lengthy, but offer no difficulties. For

$$\triangle A_s = 0\cdot004\ A_s + 0\cdot001\ A_{s\ max}, \text{ for } |\triangle Z_s''| = \tfrac{1}{16} \text{ in.},$$

and under the assumption that the absolute values of all partial errors are to be added, one obtains $|\triangle Z_{r1}| < 0\cdot0056$, $|\triangle Z_{r2}| < 0\cdot067$, $|\triangle Z_{r3}| < 0\cdot10$. Some error $|\triangle Z_{r4}|$ must be added to account for the inaccuracy in determining the position of the probe assembly. If $|\triangle Z_{r4}|$ in the Z''-plane is $\pm\frac{1}{8}$ in., then $|\triangle Z_{r4}| = 0\cdot038$. The sum of all errors is $0\cdot21$, but the probable mean error will be only a fraction of this value. $|\triangle Z_{r1}|$ is so small that it will be necessary to reduce the sensitivity of the detecting device; otherwise the irregular deviations of the probe assembly when being rotated will disturb the measurements.

TABLE 3

S	1	2	3	4	Return electrode
Z'_s	-1	$-0\cdot5$	$+0\cdot5$	$+1$	—
Z_s	$-0\cdot9+j0\cdot7$	$-0\cdot75+j0\cdot7$	$-0\cdot45+j0\cdot7$	$-0\cdot30+j0\cdot7$	—
$G(Z_s)$	$-0\cdot406+j1\cdot358$	$-0\cdot319+j0\cdot838$	$-0\cdot429+j0\cdot082$	$-0\cdot586-j0\cdot154$	
A_s	$+0\cdot271-j0\cdot905$	$-0\cdot425+j1\cdot117$	$+0\cdot572-j0\cdot1093$	$-0\cdot391-j0\cdot0127$	$-\Sigma A_s=-0\cdot027$
$990A_s$	$+268-j896$	$-421+j1106$	$+566-j108\cdot3$	$-387-j101\cdot7$	-26
Connection of: $(1+j)1000g$	-100 V.	$+100$ V.	0 V.	0 V.	
$(1+j)\ 100g$	$+100$ V.	$+100$ V.	-100 V.	-100 V.	
$(1+j)\ 10g$	$+40$ V.	$+60$ V.	-80 V.	-20 V.	
$i^{(i)}_{s0}/100g$	$-j896$	$+j1106$	$-j108$	$-j102$	0
Part $i^{(r)'}_{s0}$ of $i^{(r)}_{s0}/100g$*	-896	$+1106$	-108	-102	0
Required value $i^{(r)}_{s0}/100g$	$+268$	-421	$+566$	-387	-26
Part $i^{(r)''}_{s0}$ of $i^{(r)}_{s0}/100g$	$+1164=2\times582$	$-1527=2\times763\cdot5$	$674=2\times337$	$-285=-2\times142\cdot5$	$-26=-2\times13$
Connection of: $2000g$	$+60$ V.	-80 V.	$+40$ V.	-20 V.	0 V.
$200g$	-20 V.	$+40$ V.	-60 V.	$+60$ V.	-20 V.
$20g$	$+20$ V.	-40 V.	-20 V.	-20 V.	$+60$ V.
$i^{(r)''}_{s0}/100g$	$+1164=2\times582$	$-1528=-2\times764$	$676=2\times338$	$-284=-2\times142$	$-28=-2\times14$

* Supplied by series combination of condenser and resistor.

For the second measurements ($Z_0=-0\cdot6+j0\cdot7$ and $u=0\cdot3$) one obtains $Z'_r=0\cdot333+j0\cdot5533$ from the known value of the root for the purpose of accuracy calculation. Since $g=\frac{1}{6\times10^6}$ mhos, $CK=I_{s0}/A_s=990\times100g=16\cdot5\times10^{-3}$. K is slightly different from its first value $1\cdot017$, but the change is irrelevant. One can assume $C=16\cdot2\times10^{-3}$.

As to the components of the coefficients A_s and currents I_{s0} respectively, let $a_{s0}{}^{(i)}$ and $i_{s0}{}^{(i)}$ be the imaginary, and $a_{s0}{}^{(r)'}$ and $i_{s0}{}^{(r)'}$ the real components produced by the series combination of resistors and condensers, and $a_{s0}{}^{(r)''}$ and $i_{s0}{}^{(r)''}$ the real components produced by the resistors. $a_{s0}{}^{(r)}=a_{s0}{}^{(r)'}+a_{s0}{}^{(r)''}$. Considering the same sources of errors as before and in addition the influence of a deviation of the frequency, one obtains

$$\begin{aligned}\triangle a_{s0}{}^{(r)'} &=0\cdot014\ a_{s0}{}^{(r)'}+0\cdot001\ a_{s0}{}^{(r)}{}_{max},\\ \triangle a_{s0}{}^{(r)''} &=0\cdot004\ a_{s0}{}^{(r)'},\\ \triangle a_{s0}{}^{(i)} &=0\cdot009\ a_{s0}{}^{(i)}+0\cdot001\ a_{s0}{}^{(i)}{}_{max}.\end{aligned}$$

The calculations give

$$\begin{aligned}|\triangle Z_{r1}| &<0\cdot00027\\ |\triangle Z_{r2}| &<0\cdot020\\ |\triangle Z_{r3}| &<0\cdot005\\ |\triangle Z_{r4}| &<0\cdot005\\ \hline \text{Total} \quad &<0\cdot030\end{aligned}$$

The accuracy is approximately seven times as great as in the first measurements.

Actual measurements in the tank made with existing equipment, and representing the assumptions and conditions of the example, but with lesser accuracy in the voltages and admittances than assumed in the example, gave for the first approximate measurement with $Z_0=0$ and $u=4$ a value of $-0\cdot6+j1\cdot04$ for the root. A second measurement with a shift of origin and a change of scale, i.e. $Z_0=-0\cdot6+j0\cdot7$ and $u=0\cdot3$, gave a value for $Z'_r=0\cdot38+j0\cdot49$, from which follows $Z_r=-0\cdot486+j0\cdot847$. The total error equals $\sqrt{0\cdot014^2+0\cdot019^2}=0\cdot024$. The fact that this is less than the error computed for equipment of higher accuracy is due to partial cancellation of the various errors.

IX. FURTHER MEASUREMENTS

Further informative measurements carried out in a tank of the dimensions given in Section VII and with equipment similar to that referred to there confirmed the practicability of the methods described in this paper. It was ascertained that :

(i) The way of generating the electrode currents is satisfactory.

(ii) The method of detection and all relevant details are sound and give the result expected.

(iii) The magnitude of the errors to be envisaged complies with the theories developed in Section VI.

X. OTHER POSSIBLE USES OF THE METHODS DESCRIBED

(a) *Representation of Rational Functions with Simple Poles*

As other authors pointed out previously (Hansen and Lundstrom 1945 ; Boothroyd, Cherry, and Makar 1949) the electrolytic tank is a useful device for the determination of the steady-state frequency characteristics of linear networks of lumped elements. This determination, which is made for $Z'=Z$, can be considered a special case of the problem to find for any complex Z the value of

$$H(Z)=\frac{G(Z)}{F(Z)}, \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (48)$$

where $G(Z)$ and $F(Z)$ are given polynomials in Z and the degree of $F(Z)$ is higher than the degree of $G(Z)$.

The authors suggest the following procedure, which is different from that recommended in the above papers : After the roots Z_s of $F(Z)=0$ have been found, $H(Z)$ can by equation (**5**) be expressed as the sum of partial fractions, provided that there are no multiple roots, and a tank can be set up in which currents according to equations (**6**), (**7**), and (**12**) are impressed at the points Z_s. Let the quadruple probe assembly be placed at an arbitrary point Z of the tank. To make the field at this point circularly polarized, another additional electrode with the impressed current $I_q=CA_q$ must be placed at another point Z_q of the tank and an additional current $(-I_q)$ must be impressed at the electrode representing the point ∞. Then one obtains

$$H(Z)=\frac{A_q}{Z_q-Z}, \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (49)$$

a formula which is extremely simple. The technique of the measurements is also very simple. One is at liberty to vary either A_q or Z_q or both. The authors suggest selecting a value for A_q and varying Z_q, i.e. moving the additional electrode across the tank until its correct position is found.

The procedure appears to be simpler and not less accurate than the technique used so far by which the modulus (magnitude) of $H(Z)$ is found as a potential difference between two non-adjacent points of the tank, and the argument (phase angle) of $H(Z)$ as a flux between these points, i.e. as the integral along a straight line of the component of the voltage gradient perpendicular to this line.

The circular tank is unsuitable for the procedure suggested by the authors, because the wall of the tank would have to be insulating for the real parts of a conjugate complex pair of coefficients A_s and conducting for their imaginary parts.

(b) *Approximate Representation of Rational Functions with Multiple Poles*

Multiple roots of $F(Z)$ can be dealt with by the following approximations :

For a double root two partial fractions $\frac{1}{2}\,\frac{A_2-A_1}{Z-(Z_0-U)}$ and $\frac{1}{2}\,\frac{A_2+A_1}{Z-(Z_0+U)}$

are assumed where U is arbitrary, but $|U|$ is small compared with the distance $|Z-Z_0|$, and large compared with the diameters of the electrodes. The identity

$$P_2(Z)\cdot\left[\tfrac{1}{2}\cdot\frac{A_2-A_1}{Z-(Z_0-U)}+\tfrac{1}{2}\cdot\frac{A_2+A_1}{Z-(Z_0+U)}\right]=\frac{UA_1}{(Z-Z_0)^2}+\frac{A_2}{Z-Z_0}, \qquad \ldots\ldots\ldots\ldots (50)$$

where $P_2(Z)$ is a correction factor of the value

$$P_2(Z)=1-\left(\frac{U}{Z-Z_0}\right)^2, \qquad \ldots\ldots\ldots\ldots (51)$$

shows how to select the currents at the electrodes placed at (Z_0-U) and (Z_0+U). If the direction of U is identical with the direction of $(Z-Z_0)$ or perpendicular to it, $P_2(Z)$ is real.

A triple root at Z_0 can be approximated by three partial fractions $\tfrac{1}{3}\cdot\dfrac{A_1+A_2+A_3}{Z-(Z_0+U)}$, $\tfrac{1}{3}\cdot\dfrac{NA_1+N^2A_2+A_3}{Z-(Z_0+NU)}$, and $\tfrac{1}{3}\cdot\dfrac{N^2A_1+NA_2+A_3}{Z-(Z_0+N^2U)}$, where the same conditions for U apply as before and $N=\varepsilon^{j(2\pi/3)}=-\tfrac{1}{2}+\tfrac{1}{2}j\sqrt{3}$. The identity

$$P_3(Z)\cdot\left[\tfrac{1}{3}\cdot\frac{A_1+A_2+A_3}{Z-(Z_0+U)}+\tfrac{1}{3}\cdot\frac{NA_1+N^2A_2+A_3}{Z-(Z_0+NU)}+\tfrac{1}{3}\cdot\frac{N^2A_1+NA_2+A_3}{Z-(Z_0+N^2U)}\right]$$
$$=\frac{A_1U^2}{(Z-Z_0)^3}+\frac{A_2U}{(Z-Z_0)^2}+\frac{A_3}{Z-Z_0}, \qquad \ldots\ldots\ldots\ldots (52)$$

where

$$P_3(Z)=1-\left(\frac{U}{Z-Z_0}\right)^3 \qquad \ldots\ldots\ldots\ldots (53)$$

settles the problem for a triple root. If the direction of $(Z-Z_0)$ is the same as that of $\pm U$ or $\pm NU$ or $\pm N^2U$, $P_3(Z)$ is real.

The relation between partial fractions for multiple roots and symmetrical components of polyphase systems (Wagner and Evans 1933) exists for any order n of a root. If $M=\varepsilon^{j(2\pi/n)}$ one can derive the relation

$$\sum_{s=1}^{s=n}\frac{A_{n+1-s}\,.\,U^{s-1}}{(Z-Z_0)^s}=P_n(Z)\cdot\sum_{s=1}^{s=n}\left[\frac{1}{n}\cdot\frac{\sum\limits_{r=1}^{r=n}(A_r\,.\,M^{s.r})}{Z-(Z_0+M^sU)}\right], \qquad \ldots\ldots (54)$$

where

$$P_n(Z)=1-\left(\frac{U}{Z-Z_0}\right)^n. \qquad \ldots\ldots\ldots\ldots (55)$$

Roots of higher order than 3 are unlikely to occur in practice.

XI. ACKNOWLEDGMENT

The authors are indebted to Professor D. M. Myers, Department of Electrical Engineering, University of Sydney, for his encouragement of, and interest in, the work, and for permission to make use of the facilities of the Department for experimental work.

XII. References

Bloch, A. (1948).—Solution of algebraic equations by means of an electrolytic tank. VIIth Internat. Congr. Appl. Mech., London. Pap. No. IV–28.

Boothroyd, A. R., Cherry, E. C., and Makar, R. (1949).—An electrolytic tank for the measurement of steady-state response, transient response, and allied properties of networks. *Proc. Instn. Elect. Engrs.* **96**: 163.

Hansen, W. W., and Lundstrom, O. C. (1945).—Experimental determination of impedance functions by the use of an electrolytic tank. *Proc. Inst. Radio Engrs. N.Y.* **33**: 528.

Lucas, F. (1888).—Résolution des équations par l'électricité. *C.R. Acad. Sci. Paris* **106**: 645, 1072.

Wagner, C. F., and Evans, R. P. (1933).—"Symmetrical Components." (McGraw-Hill Book Co. Inc.: New York.)

Appendix I

Treatment of the Case where Some of the Points Z_s Correspond to Roots of the Equation

Let r points $(Z_1', Z_2', \ldots Z_r')$ of the selected m points $Z_1', Z_2', \ldots Z_m'$ $(r<m)$ correspond to roots of the equation $G(Z)=0$. This becomes manifest when the values A_s are computed, because A_s will be zero for $1<s<r$. One can reduce the degree of the equation by dividing $G(Z)$ by $\prod\limits_{q=1}^{q=r}(Z-Z_q)$ and start anew with an equation $G'(Z)=0$ of the degree $(m-r)$. Nevertheless, it is possible to ignore the fact that some of the selected points are roots of $G(Z)=0$, and the original procedure need not be changed on this account. To prove this let

$$G'(Z)=\frac{u^r \, . \, G(Z)}{\prod\limits_{q=1}^{q=r}(Z-Z_q)}=\sum_{q=0}^{q=m-r} C'_q \, . \, Z^q, \qquad (56)$$

$$A'_s=\frac{G'(Z_s)}{B'_s(Z'_s)}, \qquad (57)$$

where

$$B_s(Z'_s)=\prod_{q=r+1 \text{ to } s-1}^{q=s+1 \text{ to } m}(Z'_s-Z'_q). \qquad (58)$$

A'_s are the quantities to be used for the probe currents according to equation (**12**), if the equation $G'(Z)=0$ is to be solved and the selected $(m-r)$ points for the probes are identical with those $(m-r)$ points of the original scheme for which $G(Z_s)\neq 0$. By multiplying numerator and denominator of (**57**) by $\prod\limits_{q=1}^{q=r}(Z'_s-Z'_q)$ one obtains

$$A'_s=\frac{G(Z_s)}{B_s(Z'_s)}. \qquad (59)$$

Hence A'_s is identical with A_s defined by equation (**6**).

This investigation shows also that it is irrelevant if one of the selected points Z_s happens to be very close to a root of the original equation $G(Z_s)=0$.

Appendix II

Derivation of Equation (**34**) *for the Output Voltage of the Differential Network*

Figure 12 shows the circuit of the differential network if equations (**32**) and (**33**) are complied with. Let I_1 and I_2 be the currents in the branch *a–8–b*, if the generator V_6 or V_7 alone is active, and the other generator not working, respectively. These currents can be computed with the help of Thevenin's theorem (see Fig. 13), and one obtains

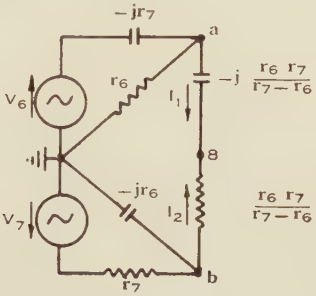

−jr7
a
r6
−j r6 r7 / r7−r6
V6
I1
8
−jr6
I2
V7
r6 r7 / r7−r6
r7
b

Fig. 12.—Equivalent circuit of the differential network.

$$I_1 = -V_6 \frac{(r_7 - jr_6)(r_7 - r_6)}{2(1+j)r_7^3}, \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{60})$$

$$I_2 = jV_7 \frac{(r_6 - jr_7)(r_7 - r_6)}{2(1+j)r_7^3}. \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{61})$$

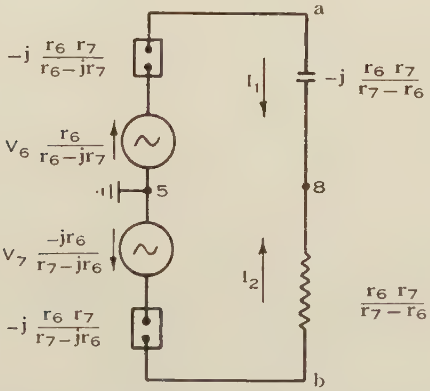

a
−j r6 r7 / r6−jr7
I1
−j r6 r7 / r7−r6
V6 r6 / r6−jr7
5
8
V7 −jr6 / r7−jr6
I2
r6 r7 / r7−r6
−j r6 r7 / r7−jr6
b

Fig. 13.—Application of Thevenin's theorem to Figure 12.

If the potentials of the point 8 caused by the currents are computed as the voltage drops in the impedances 8–*b*–5 and 8–*a*–5 respectively and superimposed upon each other, one obtains equation (34).

GALACTIC RADIATION AT 18·3 Mc/s.

By C. A. Shain*

[*Manuscript received March* 27, 1951]

Summary

Observations are described of galactic radiation at a frequency of 18·3 Mc/s. over a zone of the sky centred on Declination —34°. The results, expressed as equivalent temperatures, are compared with the distribution of intensity at 100 Mc/s. It is found that the temperatures are consistently much higher at 18·3 Mc/s. than at 100 Mc/s. but the observed ratio of maximum to minimum at 18·3 Mc/s. is less than would have been obtained at 100 Mc/s. using an aerial with the same directional diagram. The observations are consistent with the hypothesis that contours of equal equivalent temperature over the sky are of similar shape on the two frequencies. Peaks in the intensity distribution are observed which appear to be due to discrete sources.

The observations give information concerning the attenuation of radiation passing through the ionosphere. It is found that *F*-region absorption is higher near the critical frequency than expected, and that *E*-region absorption is lower than expected. It is found that the collision frequency at the height of maximum ionization of the *E* region is not greater than 10^4 sec.$^{-1}$, and is probably less.

For the equipment used, the limit of sensitivity for the reception of radio signals was generally set by galactic noise.

I. INTRODUCTION

It has been shown by several authors, particularly Townes (1947) and Piddington (in press), that observations of galactic radiation at low frequencies are of importance in relation to theories of the origin of the radiation. At low frequencies, however, two difficulties are encountered. Firstly, it is difficult to survey the sky, owing to the large size of highly directive aerials, and secondly, a considerable amount of deviation and absorption may occur in the ionosphere, particularly at low angles of elevation. The most useful series of observations has been from the early work of Jansky (1937) on frequencies near 20 Mc/s.

The present paper describes observations on a frequency of 18·3 Mc/s. with an aerial of high directivity. The latitude of the observing site was 34 °S. and the direction of maximum sensitivity of the aerial was fixed vertically upwards. The observations therefore give a weighted average of the radio emission from a zone of the sky about Declination —34°. This zone includes both the centre and the south pole of the Galaxy.

It was expected that the radiation would be attenuated in passing through the ionosphere, at least during daylight hours. Observations were therefore taken over a period of approximately six months, so that all parts of the zone could be observed at night.

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

Observations with a single fixed aerial cannot give the surface distribution of the intensity of galactic radiation over the celestial sphere. However, the observed results can be compared with those to be expected from surveys carried out on other frequencies. The results of these surveys are in the form of contours, and from comparison with the present observations, the probable arrangement of the contours for 18·3 Mc/s. has been found. The survey carried out by Bolton and Westfold (1950) on 100 Mc/s. has been used, as their survey completely covers the zone of the sky of the present observations.

II. Equipment and Observational Procedure

The aerial was an array of eight half-wave dipoles connected in phase, supported 0·2 wavelengths above ground, and arranged in plan as shown in Figure 1 (*a*). Contours of equal power sensitivity calculated from the aerial characteristics are shown in Figure 1 (*b*). The receiving site, at Hornsby, N.S.W. (lat. 33·7 °S., long. 151·1 °E.), was flat but surrounded at a distance of about 1 km. by hills which subtended an angle of elevation of approximately 9°. A standard communications receiver with a bandwidth of 400 c/s. was used. The output was indicated by a milliammeter with D.C. amplifier connected to the second detector. Observations were made manually, by substituting for the aerial a diode noise generator of the same output impedance, and adjusting the diode plate current to give the same meter reading as was obtained when the aerial was connected. A single reading could be reproduced to better than ±5 per cent.

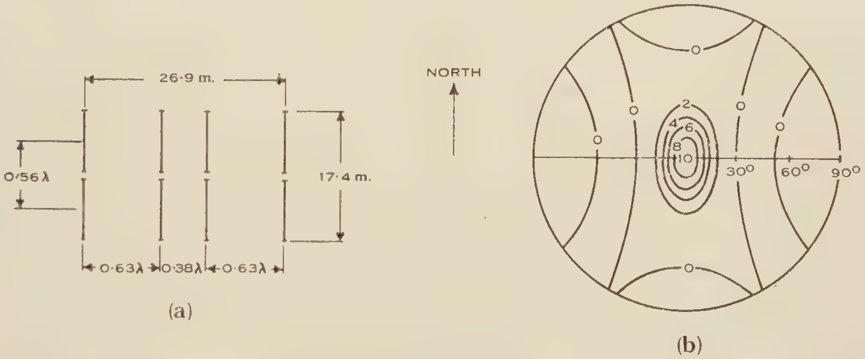


Fig. 1 (*a*).—Plan of aerial.

Fig. 1 (*b*).—Calculated contours of equal power sensitivity.

Before each reading was taken, the receiver was tuned to a frequency at which no interfering stations could be heard. This was always possible in the range 18·28 to 18·32 Mc/s. Bursts of solar noise were observed but these could be recognized, and caused no interference to the observations. It was in fact found that except for local thunderstorm activity, the limit to the sensitivity of the receiving equipment was always set by galactic noise.

Observations were made at intervals from May to November 1949. The readings of noise-generator plate current were converted to equivalent temperature, and allowance was made for losses in the aerial feeders and earth. There is an uncertainty of about 20 per cent. in the estimate of feeder losses.

The corrected equivalent aerial temperatures are plotted against sidereal time in Figure 2.

III. Ionospheric Effects

Part at least of the considerable scatter in equivalent temperatures was thought to be due to the ionosphere. The observations were therefore grouped in time from half an hour before to half an hour after each sidereal hour ; then, for each group, the observed equivalent temperatures were plotted, as a scatter diagram, against the critical frequency of the F_2 region, f^0F_2. Hourly values were obtained from the ionospheric recording station at Canberra (35·3 °S., 149·0 °E.).

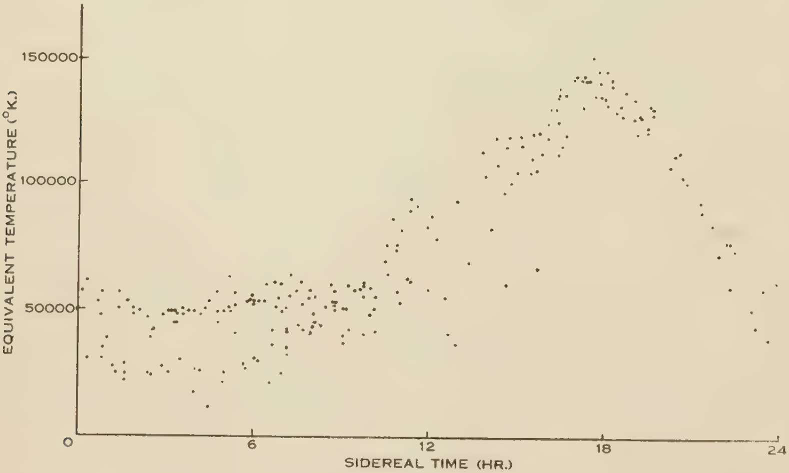


Fig. 2.—Equivalent aerial temperature *v*. sidereal time (all observations).

It was found that for f^0F_2 at Canberra greater than about 9·0 Mc/s., the equivalent temperatures were consistently low, while for f^0F_2 less than 9·0 Mc/s. they were, within about ±10 per cent., independent of f^0F_2.

To illustrate this, the results of all sidereal times have been collected together in the following manner. In each hourly group those observations for which f^0F_2 at Canberra was less than 9·0 Mc/s. were averaged. Some groups were subdivided if the rate of change of equivalent temperature was high. Then, for all observations, the ratio of equivalent temperature to the group average was plotted against f^0F_2 at Canberra. These points are shown in Figure 3, and it is evident that the measured equivalent temperatures decrease for the higher values of critical frequency.

It was expected that attenuation due to refraction and absorption would arise in the F region. Bremmer (1949) gives the theory of the attenuation due to refraction ("divergence") and it is found that for the experimental conditions the loss due to refraction effects is small, being only 0·3 db. when f^0F_2 is 12 Mc/s.

Absorption due to electronic collisions, when the operating frequency is close to the critical frequency, has been studied by Jaeger (1947) for a Chapman region. He gives the total absorption at vertical incidence for a wave of frequency f making a double passage of a Chapman region as

$$\frac{\nu_0 H f_0^{\,2}}{cf^2}(\sec\chi)^{-3/2}\Phi\left(\frac{f_c}{f}\right) \text{ nepers,} \quad \ldots\ldots\ldots\ldots\ldots \quad (1)$$

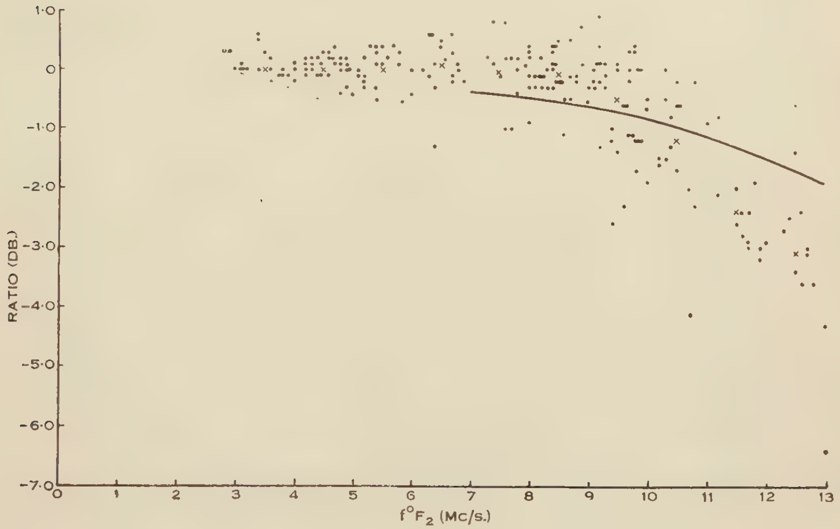


Fig. 3.—Ratio of observed intensities to hourly average with $f^0F_2 \leqslant 9\cdot 0$ Mc/s., plotted against f^0F_2. Crosses represent averages for 1 Mc/s. intervals of f^0F_2. The curve is the theoretical absorption in a Chapman region calculated according to equation (2), taking $H\nu_m = 4\times 10^4$ km./sec.

where ν_0 is the collision frequency at the height of maximum ionization and f_0 the critical frequency when χ the zenith angle of the Sun is zero,

f_c is the critical frequency at the time of observation,

H is the scale height of the atmosphere,

c is the velocity of light, and

$\Phi(f_c/f)$ is a function which is tabulated in the paper. For small values of the ratio f_c/f it is slowly varying, and the limiting value, for f_c very small compared with f, is 4·13.

For the case of the galactic noise passing once through the layer, the absorption given by (1) would be halved, and on substituting

$$\nu_m = \nu_0 \cos\chi,$$

where ν_m is the collision frequency at the height of maximum ionization for the time of observation, and putting

$$f_c^2 = f_0^2 \cos^{\frac{1}{2}} \chi,$$

the absorption of galactic noise is found to be

$$\frac{4 \cdot 34 H \nu_m}{c} \cdot \frac{f_c^2}{f^2} \cdot \Phi(f_c/f) \quad \text{db.} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

These formulae assume that the operating frequency is much greater than the collision frequency. If this is not the case, Jaeger gives a correction factor which must be applied to (1) and (2). Also, the influence of the Earth's magnetic field has been neglected. This should raise the absorption of the extraordinary ray and lower the absorption of the ordinary ray by nearly equal small amounts. Since, however, the aerial does not discriminate between the two components, and assuming that the polarization of the radiation incident on the ionosphere is random, the absorption should be given very nearly by (2).

A curve showing the absorption calculated according to (2) (taking $H\nu_m = 4 \times 10^4$ km./sec. for the F region) with the small correction for refraction effects is included in Figure 3. It is seen that the observed attenuation increases more rapidly with f^0F_2 than expected. Some part of this effect may be caused by an increased attenuation due to the Earth's magnetic field which is greater than the simple theory would suggest. However, unexpectedly large attenuation in the F region has also been found in the reception of radio echoes from the Moon (Kerr and Shain 1951). In that case it was associated with excessive refraction and this may be so with the present observations.

Although there is a close correlation between attenuation and f^0F_2, a check was made for absorption in lower layers. For this purpose those observations for which f^0F_2 was less than 9·0 Mc/s. were used. As before, the ratio of each observed temperature to the group average was taken, and in Figure 4 they are plotted against the critical frequency of the E layer, f^0E, at Canberra. The group of points to the left of the figure corresponds to values of f^0E less than 1·5 Mc/s. It is seen that there is little variation with f^0E and, in fact, for f^0E up to 3·5 Mc/s. the absorption is less than 0·2 db. at 18·3 Mc/s. Substitution of these values in (2) gives an upper limit to the product $H\nu_m$ of $0 \cdot 9 \times 10^5$ km./sec., at the level of maximum ionization of the E region, even if all the absorption occurs in this region. Taking the scale height as 9 km., this implies that ν_m is not greater than 10^4 sec.$^{-1}$ at this level.

Taylor (1948) has given results of observations of absorption of continuous waves at nearly vertical incidence on frequencies of 2·061 and 4·272 Mc/s. The results, which cover observations extending over several years, give the absorption of waves of these frequencies, compared with the " unabsorbed " signal strength of signals received at night, for different values of the Sun's zenith angle. The figures given are the values corrected for the effects of the Earth's magnetic field and for multiple reflections—that is, for the ordinary ray.

The figures for the absorption of the 4·272 Mc/s. radiation at noon, local time, have been used in (2), in conjunction with the published observations of median values of f^0E at Washington, to find the corresponding value of $H\nu_m$. In the substitution, the double passage of the E region was included and the frequency was taken as $(4 \cdot 272 + f_H)$ Mc/s., where f_H is the gyro-frequency, since Taylor's results are given for the ordinary ray.

The mean value obtained for $H\nu_m$ is $3 \cdot 5 \times 10^5$ km./sec. and a curve of absorption expected using this value of $H\nu_m$ is shown in Figure 4. The value of $H\nu_m$ deduced from Taylor's results agrees with the results obtained by means of pulse observations (e.g. Best and Ratcliffe 1933); it is, however, somewhat higher than the results obtained in the present work. This may be due to the different circumstances of the two types of experiment. Firstly, measurements on waves reflected from the F region must necessarily be carried out at a comparatively low frequency, and secondly, the received signal shows a considerable degree of fading, whereas the galactic noise observations use a small band of frequencies and the received signal, which in the meter circuit is averaged over a time somewhat less than one second, shows only small fluctuations.

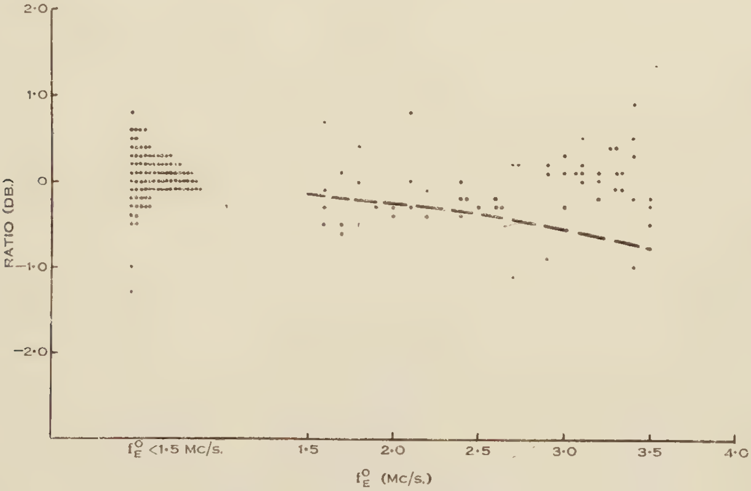


Fig. 4.—Ratio of observed intensities ($f^0F_2 \leqslant 9 \cdot 0$ Mc/s.) to hourly averages plotted against f^0E. The curve is the theoretical absorption in a Chapman region, calculated according to equation (2), taking $H\nu_m = 3 \cdot 5 \times 10^5$ km./sec.

It should be pointed out that all the above measurements of lower-layer absorption make no distinction between absorption in the E region and the D region. Appleton and Piggott (1948), by comparing the absorption of waves reflected from the E and F regions, have shown that the region mainly responsible for the absorption is situated lower than, and distinct from, the E region. If that is so, the upper limit to $H\nu_m$ at the level of maximum ionization of the E region, $0 \cdot 9 \times 10^5$ km./sec., must be considerably reduced.

Finally, Taylor has found a variation of the absorption with sunspot number which is probably connected with the normal variation of f^0E with sunspot number. No significant correlation of absorption with sunspot area or sunspot number has been found in the present limited series of observations.

From the foregoing analysis it was considered that all observations for which f^0F_2 was less than 9·0 Mc/s. could be taken as unaffected by the ionosphere. These observations are shown individually in Figure 5 and in Figure 6 (*a*) as a curve through the hourly averages. Relative intensities should be correct to better than ±10 per cent., but the absolute values may be 10 per cent. high or 20 per cent. low, mainly owing to uncertain losses in the aerial system.

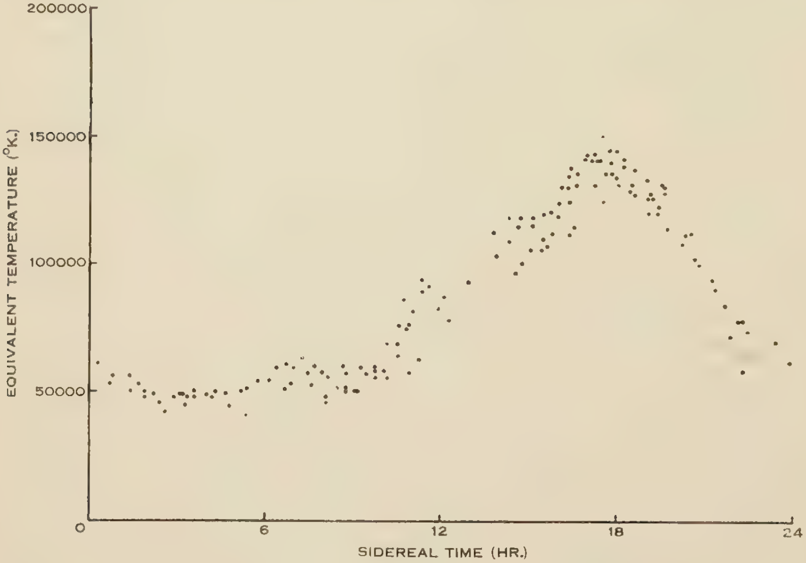


Fig. 5.—Observations used in analysis. Equivalent temperature *v.* sidereal time ($f^0F_2 \leqslant 9\cdot0$ Mc/s.).

IV. Comparison with 100 Mc/s. Survey

Because the observations were limited to a fixed aerial direction, and since the aerial has an angular beam width comparable with the dimensions of certain features of the galactic noise distribution, the result of the observations (Fig. 6 (*a*)) gives only a rough indication of the noise distribution across the sky. As a comparison with work on other frequencies, the equivalent aerial temperatures to be expected from the contours of Bolton and Westfold (1950) were calculated. The aerial diagram (Fig. 1 (*b*)) was divided into small sections, and for each hour of Right Ascension (Dec. −34°), it was superimposed on the 100 Mc/s. contours. The weighted average of equivalent temperatures in the aerial beam was found, and, for comparison with the present observations, the temperatures were multiplied by a constant factor, so that the calculated and observed temperatures agreed in the region of the galactic pole. The curve of calculated equivalent aerial temperature (based on the 100 Mc/s. survey) against sidereal time is shown in Figure 6 (*b*). It is seen that although the shape of the two curves is roughly similar, the observed ratio of maximum to minimum is considerably lower than that expected from 100 Mc/s. results.

Surveys of the distribution of galactic noise have been made at frequencies from 480 to 64 Mc/s. (Reber 1944, 1948; Hey, Parsons, and Phillips 1948; Allen and Gum 1950; Bolton and Westfold 1950) and comparison shows that the shapes of contours of equal noise intensity are broadly similar, although the ratio of maximum to minimum intensity may decrease with frequency. To obtain the probable distribution of intensity at 18·3 Mc/s., the assumption was made that the similarity in shape of the contours holds down to this frequency.

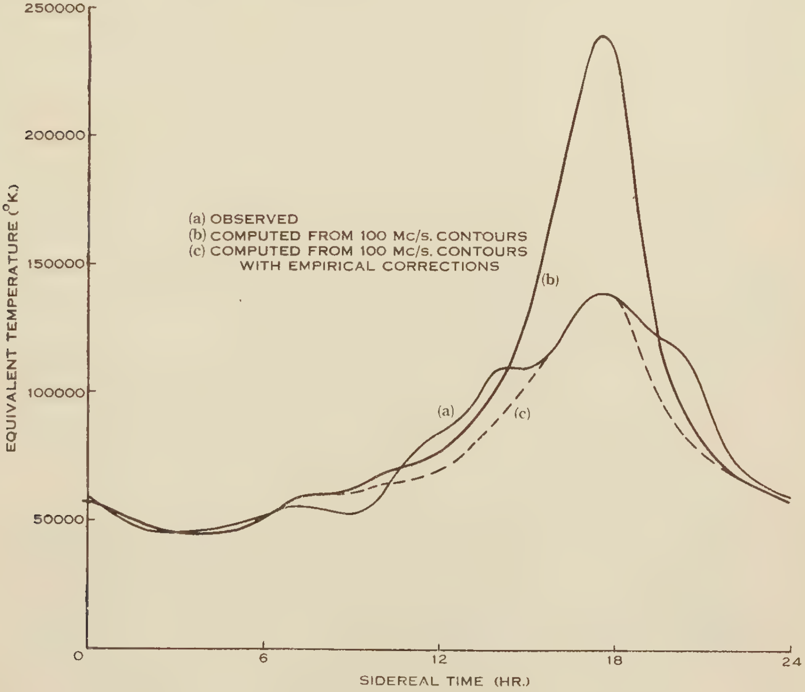


Fig. 6.—Equivalent aerial temperature *v.* sidereal time.

Then a map was taken of the 100 Mc/s. contours and temperatures were attached to them, by trial and error, so that when weighted averages over the aerial diagram were taken as before, the computed and experimental curves agreed.

It was found that a good fit could be obtained by using the relation

$$T_{18\cdot3}=200{,}000(1-e^{-T_{100}/1900})\ ^{\circ}\text{K.},$$

where $T_{18\cdot3}$ is the temperature to be attached to the contour corresponding to T_{100}, in the 100 Mc/s. survey.

The curve computed from the result is shown in Figure 6 (*c*).

It is of interest that, to the extent to which the assumption of similarity of contours is correct, the observations of galactic noise on 100, 64, and 18·3 Mc/s. are all consistent with a simple extension of the above relation—namely

$$T_f=200{,}000(1-e^{-144Rf^{-2\cdot73}})\ ^{\circ}\text{K.},$$

where T_f is the temperature, at frequency f Mc/s., to be attached to the contour which is given the value R in Figure 8 of Bolton and Westfold's (1950) paper.

It is seen in Figure 6 that the difference between curves (*a*) and (*c*) is greatest between about 11 hr. to 15 hr. sidereal time and 19 hr. to 22 hr. sidereal time. These differences cannot be reduced by simple renumbering of the contours. One possible explanation of at least part of the discrepancy is the effects of radiation from discrete sources of noise. Stanley and Slee (1950) give a list of some discrete sources observable in latitude 34 °S. and, for some of these, the variation of intensity of radiation with frequency from 160 to 60 Mc/s. One of the most important, Centaurus-A (R.A. 13 hr. 22 min., Dec. −42·6°), passes through the aerial beam near the direction of maximum sensitivity (see Fig. 7) and, by extrapolation of the frequency spectrum of Stanley and Slee, should increase the equivalent aerial temperature at transit by approximately 6000 °K.

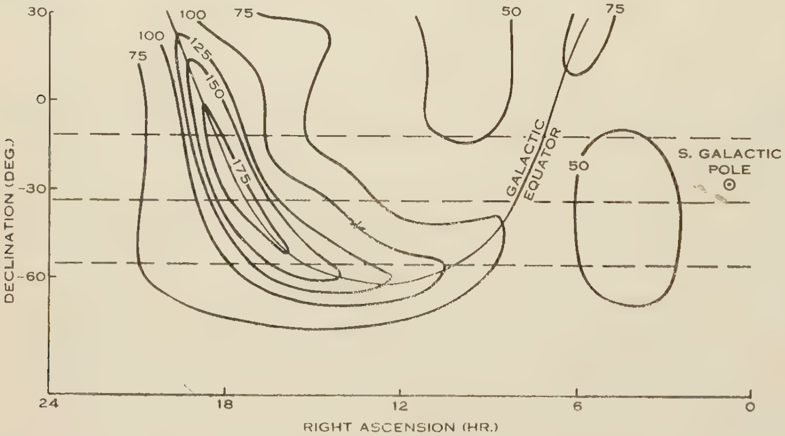


Fig. 7.—Estimated contours of equal equivalent temperature at 18·3 Mc/s. (thousands of degrees). The broken lines show the lines traced out by the centre of the aerial beam and half-power points. *X* is the position of source Centaurus-A.

The source in Cygnus (R.A. 19 hr. 58 min., Dec. +40·6°) should be six times as intense as Centaurus-A. However, the zenith angle at transit is 74°, so that the gain of the aerial in this direction should be low ; but at such a low angle of elevation there is uncertainty concerning the effect of the surrounding hills and deviation in the ionosphere. It is therefore probable that radiation from this source could account for some of the discrepancy near 20 hr. sidereal time, although it is unlikely that it would account for all. The possibility that the high values at particular sidereal times could be caused by unexpectedly large side lobes in the direction of the galactic centre at those times can probably be eliminated, as the aerial was symmetrical and the site reasonably so, and there is no symmetry in the observed curve about 17 hr. 30 min. sidereal time.

It therefore appears that some of the divergence between curves (*a*) and (*c*) of Figure 7 is due to the presence of discrete sources and that these sources are more prominent relative to the general background at low frequencies, as would be expected from the spectra given by Stanley and Slee for the respective sources.

The contours obtained as above, and corresponding to the curve Figure 6 (*c*) are shown in Figure 7. They fit the present observations to within 1 db.

V. Conclusions

These observations confirm that the intensity of galactic radiation at 18·3 Mc/s. is high and of the order of magnitude suggested by the observations of Jansky. The high level of galactic noise generally limits the sensitivity of receiving equipment at this frequency, and therefore a study of its variation with time and ionospheric conditions is of importance for telecommunication practice.

The observations indicate anomalies in the attenuation of radio waves penetrating the ionosphere, F-region attenuation being higher and lower-layer attenuation being lower than expected. In particular, the collision frequency at the height of maximum ionization of the E region is not greater than 10^4 sec.$^{-1}$ and is probably considerably less.

The observations are reasonably consistent with the hypothesis that contours of equal intensity of galactic radiation have the same shape at 18·3 Mc/s. as at 100 Mc/s. Certain discrepancies indicate that at least some discrete sources radiate at this frequency, the intensity being approximately that expected from the results of Stanley and Slee.

VI. Acknowledgments

The author wishes to thank Dr. J. L. Pawsey, Dr. J. H. Piddington, and Dr. R. N. Bracewell, Division of Radiophysics, C.S.I.R.O., for helpful discussions in the course of preparation of this paper.

VII. References

Allen, C. W., and Gum, C. S. (1950).—*Aust. J. Sci. Res.* A **3** : 224-33.

Appleton, E. V., and Piggott, W. R. (1948).—U.R.S.I. Proceedings of General Assembly, Stockholm, July 1948, 320-2.

Best, J. E., and Ratcliffe, J. A. (1938).—*Proc. Phys. Soc.* **50** : 233-46.

Bolton, J. G., and Westfold, K. C. (1950).—*Aust. J. Sci. Res.* A **3** : 19-33.

Bremmer, H. (1949).—" Terrestrial Radio-Waves." (Elsevier Publishing Co. Inc. : New York.)

Hey, J. S., Parsons, S. J., and Phillips, J. W. (1948).—*Proc. Roy. Soc.* A **192** : 425-45.

Jaeger, J. C. (1947).—*Proc. Phys. Soc.* **59** : 87-96.

Jansky, K. G. (1937).—*Proc. Inst. Radio Engrs. N.Y.* **25** : 1517-30.

Kerr, F. J., and Shain, C. A. (1951).—*Proc. Inst. Radio Engrs. N.Y.* **39** : 230-42.

Piddington, J. H.—The origin of galactic radio-frequency radiation. *Mon. Not. R. Astr. Soc.* (in press).

Reber, G. (1944).—*Astrophys. J.* **100** : 279-87.

Reber, G. (1948).—*Proc. Instn. Radio Engrs. N.Y.* **36** : 1215-8.

Stanley, G. J., and Slee, O. B. (1950).—*Aust. J. Sci. Res.* A **3** : 234-50.

Taylor, E. W. (1948).—*Bur. Stand. J. Res. Wash.* **41** : 575-9 ; Res. Pap. RP 1939.

Townes, C. H. (1947).—*Astrophys. J.* **105** : 235-40.

ANGULAR DISTRIBUTION OF α-PARTICLES AND PROTONS FROM Li^6+D^2

By D. N. F. Dunbar* and F. Hirst*

[*Manuscript received March* 27, 1951]

Summary

The angular distribution of α-particles from $Li^6(d,\alpha)He^4$ has been investigated for deuteron energies between 200 keV. and 1 MeV. The distribution may be represented by the expression

$$N(\theta)=N(90°)[1+A(E)\cos^2\theta],$$

where $A(E)$ is a function of deuteron energy. The variation of $A(E)$ with energy is due to the increasing penetration of the potential barrier by d deuterons.

The angular distributions of the two groups of protons from $Li^6(d,p)Li^{7*}$, Li^7 have been investigated in the same energy range. The distributions are complex and require d deuterons for their complete specification in the form

$$N(\theta)=a_0P_0+a_1P_1+a_2P_2+a_3P_3+a_4P_4,$$

where P_n is the Legendre polynomial of order n.

The results appear to resolve some differences apparent in earlier work. There is no evidence that the short range proton distribution is more complex than that of the long range group.

I. Introduction

Three of the reactions which arise from the deuteron bombardment of Li^6 are :

$$Li^6+D^2 \rightarrow 2He^4+22\cdot23 \text{ MeV.} \quad \ldots\ldots\ldots\ldots\ldots \quad (1)$$

$$Li^6+D^2 \rightarrow Li^7 \ +H^1+5\cdot006 \text{ MeV.} \quad \ldots\ldots\ldots\ldots \quad (2)$$

$$Li^6+D^2 \rightarrow Li^{7*}+H^1+4\cdot523 \text{ MeV.} \quad \ldots\ldots\ldots\ldots \quad (3)$$

These reactions are important because of the information they provide concerning excited states of Be^8 and Li^7. The angular distribution of the protons of reactions (2) and (3) should, in particular, aid the assignment of the correct spin to the 480 keV. level in Li^7.

Heydenburg *et al.* (1948) have investigated the angular distribution of the α-particles from reaction (1) between 0·5 and 3·5 MeV. using a proportional counter. Below 1 MeV. they find that the distribution may be represented by

$$N(\theta)=N(90°)[1+A(E)\cos^2\theta].$$

Above approximately 1·3 MeV., however, an additional term $B(E)\cos^4\theta$ is required to describe the distribution. The values of $A(E)$ are small, rising from zero at zero bombarding energy to 0·5 near 1·0 MeV. and reaching a maximum of 0·92 at about 2·0 MeV. $B(E)$ is also positive and rises from zero at 1·3 MeV. to a constant value of 0·34 above 2 MeV.

---

* Physics Department, University of Melbourne.

The variation of $A(E)$ and $B(E)$ with energy has been accounted for theoretically by Resnick and Inglis (1949). They show that the increase of both $A(E)$ and $B(E)$ from zero is due to the increasing penetration of the potential barrier by the d deuterons.

The proton distribution has been determined by Whaling and Bonner (1950) and more recently by Krone, Hanna, and Inglis (1950). The distribution is not symmetrical about the plane through the target normal to the incident beam and is much more complex than the α-particle distribution. The two sets of results are in serious disagreement, however, since Krone, Hanna, and Inglis do not confirm the rather complex distribution of the long range protons discovered by Whaling and Bonner. It is important to resolve this difference since the assignment of spin to the levels of Li^7 is dependent upon the complexity of the distributions.

The present paper describes measurements on both the α-particle and proton distributions for deuteron energies ranging from 200 keV. to 1 MeV. The photographic plate technique has been used to record the particles over a wide range of angles, since it provides a means of observing the nature of the particle as well as its range and direction. The number of particles counted—about 20,000 at each energy for α-particles and the total proton emission—is approximately the same as in the earlier experiments (Heydenburg *et al.* 1948 ; Krone, Hanna, and Inglis 1950 ; Whaling and Bonner 1950).

II. Experimental

(a) *The Electrostatic Generator*

The 1 MV. electrostatic generator used in this work has already been described (Martin *et al.* 1949). Since then an additional charging belt has been installed giving a charging current of 450 to 500 μA. Both the generator voltage and magnet current have been stabilized by electronic devices to such a degree that the energy of the particles is maintained to approximately 2 keV. at all bombarding energies. Although the magnet current is maintained to better than 0·1 per cent., it is not employed as a direct measure of particle energy since the hysteresis of the iron core causes variations of ±0·5 per cent. in the magnet current for a given particle of energy. With care, however, it is possible to reduce this variation, in which case the value of magnet current has been most useful in detecting an error in the generating voltmeter calibration, due to the dry rectifier employed in the circuit. A 6H6G diode rectifier has been substituted for the dry rectifier with a resultant improvement in stability. The non-uniform distribution of potential down the column has resulted in the occurrence of a brush discharge at the junction of the high potential electrode and the insulator. Initially, a large insulated plastic curtain was hung from the ceiling midway between the electrode and the nearest wall of the room to prevent the discharge by changing the shape of the field. While this was partially successful, it introduced a 20 per cent. change in the rotating voltmeter calibration and was consequently dispensed with in favour of two corona rings placed round the main insulator near the high potential electrode. These have been most successful in distributing the potential along the upper section of the column and steady operation is possible at the highest potential. Furthermore, no perceptible change of voltage calibration has resulted.

(*b*) *The Target Camera and Photographic Plates*

The same target chamber camera has again been employed (Martin *et al.* 1949) with cellophane windows as before to exclude the scattered deuterons from the area to be counted. Although the plates record both the α-particles and the protons, the two groups of particles have been effectively separated by adopting different techniques in development. The α-particles, which have a shorter range than the protons, are recorded on Ilford type D.1 nuclear emulsion plates which are relatively insensitive to protons. The use of a strong caustic developer for only 3¼ minutes at 18 °C. further inhibits the development of the faint proton tracks. The appearance of a typical field of view is shown in Plate 1 in which the α-particles may be distinguished immediately from the protons.

Although independent observers have shown excellent agreement in distinguishing the α-particle tracks, it is impossible to ensure that all the tracks counted are really α-particles. Variation in development and emulsion sensitivity at different points in the plate undoubtedly introduces fluctuations in the numbers of tracks counted as α-particles.

Thicker cellophane windows which virtually eliminate the α-particles simplify the recording of the proton tracks. Ilford type C.2 plates are used with normal development in dilute D.19 developer.

The Li^6 targets used have been deposited directly on to the aluminium-leaf target blank used in the camera by a 60° sector mass spectrograph, which completely resolves the mass 6 and mass 7 peaks. The average deposit used is approximately 200 microcoulombs over an area of 3 sq. mm.—a thin target. The appearance of fewer than 1 in 1000 short α-particles on the plates indicates that the amount of Li^7 in the target is negligibly small.

(*c*) *Measurement of the Photographic Plates*

The areas to be counted are marked on the plates by scribing with a fine needle point concentric circles of radii 4·0 and 4·5 cm. with the target as centre. Radial lines are also scribed at the angles where the yield is to be measured. The angles used vary from one energy to another and are chosen to give a uniform distribution of points along the angular yield curves of Figures 2 and 4.

The tracks—either α-particles or protons—are counted using a binocular microscope with a ¼ inch objective and ×6 eyepieces. An eyepiece graticule in the form of a square facilitates the determination of the number of particles per unit area of the plate at each angle counted. No further measurements are necessary for the α-particles but for the $Li^6(d,p)$ Li^{7*}, Li^7 reaction the two proton groups must be resolved, since, in general they may be expected to have different angular distributions.

For two of the proton plates at each energy, histograms of 300 tracks are compiled at each of the five angles, 0, 55, 90, 130, and 180°, measured in the plane of the plate ; i.e. each histogram consists of 600 tracks in all. Some typical histograms are shown in Figure 1. The compilation of the histograms has been facilitated by the use of a microscope with a rotating stage whose axis of rotation coincides with the position of the target on the plate and not with the optical axis of the microscope. A radial motion of the stage permits the location of tracks on the plate in terms of polar rather than cartesian coordinates. In dividing the total numbers of protons counted into the long and short range groups, smooth curves have been drawn through the observed ratios so that interpolations may be made at each of the other angles counted. In drawing these curves account has been taken of the gradual and smooth increase in the number of short range tracks with increasing deuteron energy. The statistical fluctuations in the ratio of short to long protons determined from the histograms are thus considerably reduced.

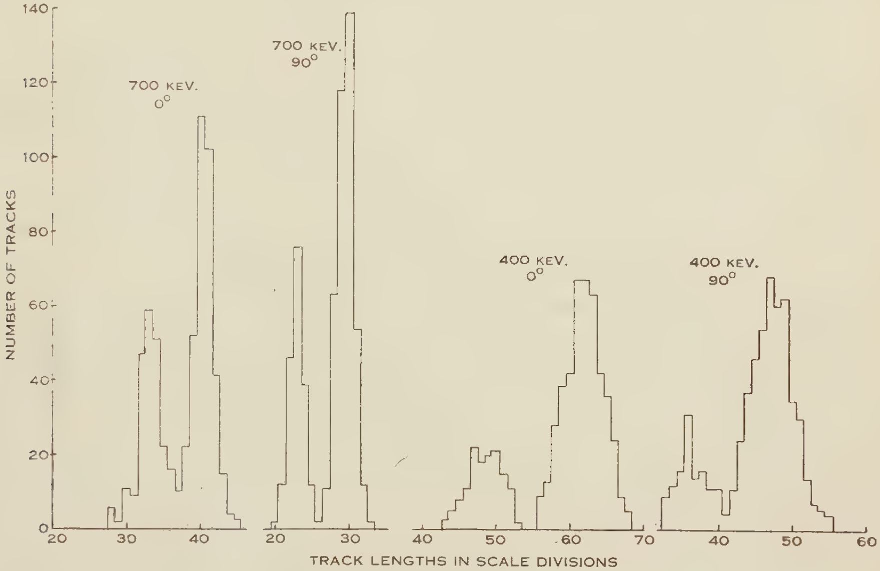


Fig. 1.—Typical histograms of protons from Li^6(*d*,*p*)Li^{7*},Li^7.

III. Results and Discussion

(*a*) *Li*^{6}(d,α)*He*^{4}

The angular distribution of the α-particles has been determined at each of the following deuteron energies : 200, 300, 400, 500, 600, 700, 800, 850, 900, 950, and 1000 keV. At least two plates have been exposed and approximately 20,000 tracks counted at each energy.

Since the target spot is not in the plane of the plate, the angles in this plane must be corrected to give the true angle to the incident beam (Martin *et al.* 1949). The angles so obtained and the solid angle subtended at the target by the area in which the tracks are counted are then converted to the centre of mass coordinate system as previously described (Martin *et al.* 1949), using the appropriate masses and Q value.

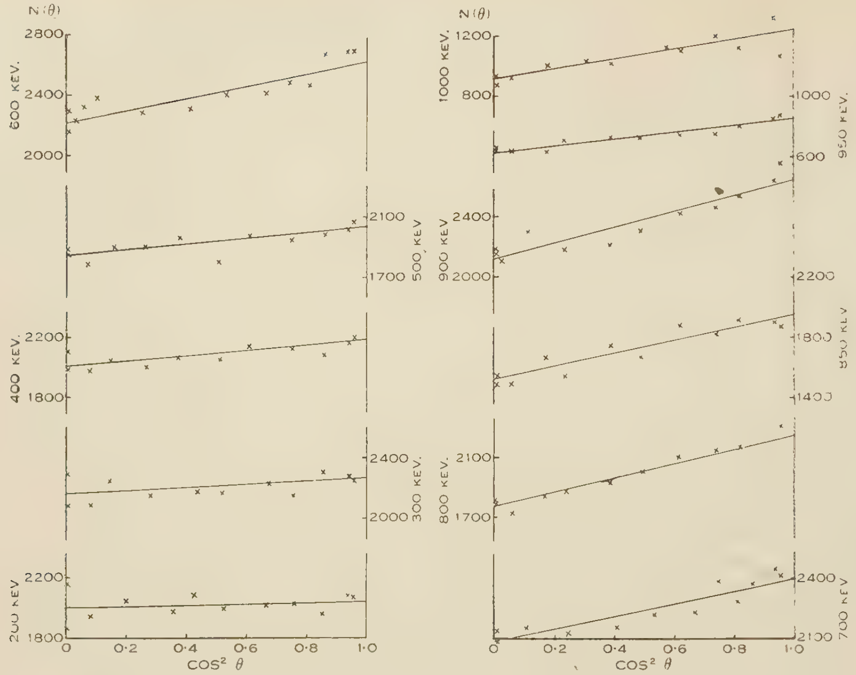


Fig. 2.—Angular distribution of α-particles from $Li^6(d,\alpha)He^4$.

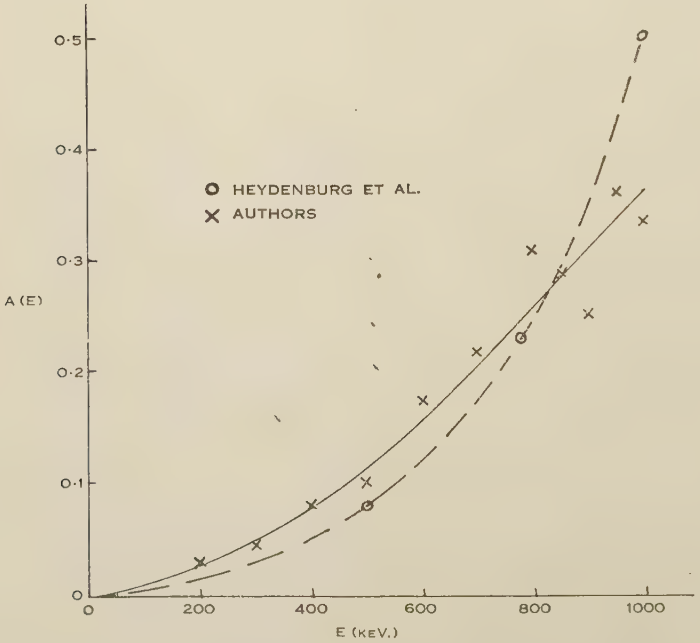


Fig. 3.—Variation of $A(E)$ with energy for $Li^6(d,\alpha)He^4$.

The relative yields $N(\theta)$ are then plotted against $\cos^2\theta$ for the various deuteron energies as shown in Figure 2. An analysis of the points using a method of curve fitting described by Guest (1950) reveals that a straight line,

$$N(\theta)=N(90°)[1+A(E)\cos^2\theta],$$

is the curve of best fit to the points.

The values of $A(E)$ are plotted against energy, E, in Figure 3. The present results are in general agreement with those of Heydenburg *et al.* (1948) which have also been included in the figure. The apparent dip in the $A(E)$ curve at 900 keV. is not considered to be real in view of the expected error in the value of $A(E)$. An analysis of the individual plates at each energy reveals a range of as much as 20 per cent. in the value of $A(E)$, although this error may be reduced in the mean to as little as 5 per cent. This view is supported by the fact that the variation of $A(E)$ with energy is due to the increasing penetrability of the d wave deuterons. The increase of $A(E)$ with energy should be gradual and would be unlikely to exhibit the fine structure which the low value of $A(E)$ at 900 keV. suggests.

In accounting for the experimental distribution observed by Heydenburg *et al.* (1948), Resnick and Inglis (1949) make the assumption that the final state of two α-particles must arise from a level in Be^{8*} of even angular momentum and even parity. Such a state can arise from the interaction of only s and d deuterons with the even Li^6 nucleus. If only s deuterons are involved, the distribution is spherically symmetric. The d wave must then be invoked to explain the observed asymmetry. By assuming the formation of two compound nuclear states of $j=0$ and $j=2$ and relating the variation of $A(E)$ to the increasing penetrability of d wave deuterons with energy, Resnick and Inglis (1949) derive an expression for the variation of $A(E)$ with energy. Since there are eleven parameters in the expression for $A(E)$, an evaluation of the resonance energy of the $j=2$ state of Be^{8*} is not possible.

There is no experimental evidence for the presence of the $B(E)\cos^4\theta$ term in the distribution below 1 MeV. Resnick and Inglis (1949) point out that the value of $A(E)$ should increase from zero, at a lower energy than $B(E)$ since $B(E)$ depends only on the square of the ratio

$$\frac{\text{Penetrability of } d \text{ wave}}{\text{Penetrability of } s \text{ wave}}.$$

The present results are in agreement with this theory over the limited energy range of the experiment.

(b) $Li^6(\text{d,p})Li^{7*}$, Li^7

The angular distribution of the protons has been determined at deuteron energies of 300, 400, 500, 600, 700, 800, and 900 keV. Some 20,000 tracks have been counted and at least two plates exposed at each energy.

(i) *Distribution of Total Proton Emission.*—This is dependent only on the total count of tracks and is statistically much more reliable than the distributions of the long and short groups which are derived from histograms. The variation of intensity with angle in the centre of mass system is shown in Figure 4, where the relative numbers of protons are plotted against cos θ. Corrections to the centre of mass system have been made using a mean Q value which introduces an error of less than 1·5 per cent.

No attempt has been made to normalize the two groups of results (0–90° and 90–180°) at 90°, in spite of the fact that at some energies there is a discrepancy in the two intensities at 90°. Since the target spot is of finite size, and the departure from uniformity over the cross section of the beam is uncertain, it seems preferable to regard the two values of the 90° yield as being of equal weight and to accept their mean as the most probable value. The uncertainty in the exact location of the most intense spot on the target will, of course, also affect other angles where correction is more difficult. The α-particle results, which are symmetrical about the plane through the target normal to the incident beam, indicate that at other angles little discrepancy exists even although the two 90° points may differ by as much as 10 per cent. The use of several plates at each bombarding energy and the statistical variation in target position also help to reduce any error. While the occasional discrepancy in the 90° yields is regarded as unsatisfactory, no convincing explanation of its origin has been found.

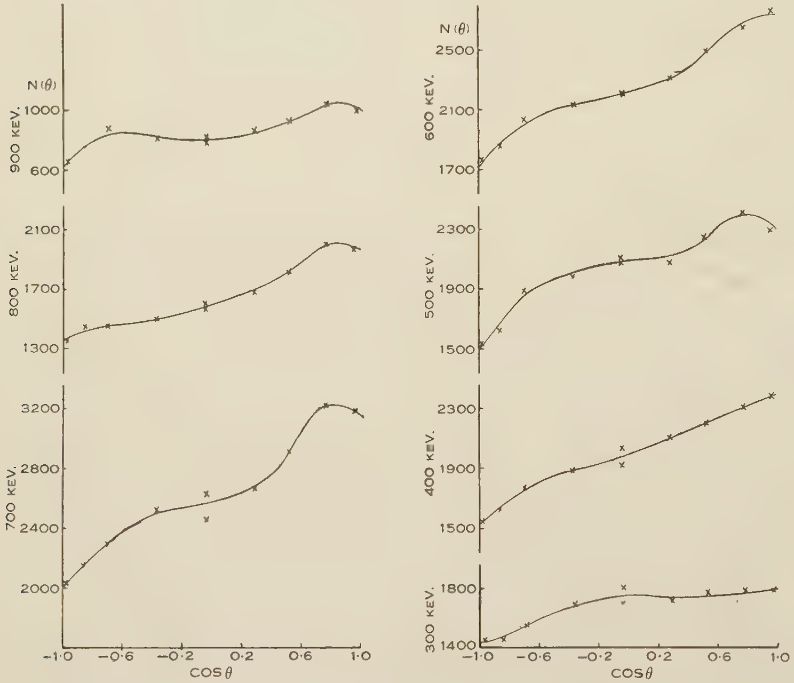


Fig. 4.—Angular distribution of total proton emission from $Li^6(d,p)Li^{7*}$, Li^7.

These results have been expressed in terms of Legendre polynomials in the form

$$N(\theta)=a_0P_0+a_1P_1+a_2P_2+a_3P_3+a_4P_4+a_5P_5+a_6P_6.$$

Such a polynomial expression is suggested since the matrix elements $H_{PR}H_{RQ}$ of the dispersion formula may be expressed in terms of Legendre polynomials.

The values of the coefficients, a_n, have been determined from

$$a_n = \frac{2n+1}{2}\int_{-1}^{+1} N(\mu) \,.\, P_n(\mu) \,.\, d\mu,$$

where $\mu = \cos\,\theta$,

$N(\mu)$ =intensity at angle θ,

$P_n(\mu)$=Legendre polynomial of order n,

by numerical integration using Simpson's rule.

The curves at various energies are then compared through the ratios a_1/a_0, a_2/a_0, a_3/a_0, a_4/a_0, etc.

The values of these ratios as a function of energy are shown in Table 1 and Figure 5 together with those of Whaling and Bonner (1950) in the same energy range. Apart from the difference in the value of a_1/a_0 at 900 keV., the two groups of results are in excellent agreement. Krone, Hanna, and Inglis (1950) determine the distributions of the proton groups separately, and the information is insufficient to find the distribution of the total emission.

TABLE 1

VALUES OF a_1/a_0, a_2/a_0, a_3/a_0,. . . .FOR TOTAL PROTON EMISSION

Energy (keV.)	a_1/a_0	a_2/a_0	a_3/a_0	a_4/a_0	a_5/a_0	a_6/a_0	Reference
300	0·097	—0·059	0·017	—0·018	—0·020	0·002	
400	0·200	—0·011	0·026	—0·018	—0·005	0·012	
	0·112	—0·002	0·034	—0·028	—0·004	—0·011	Whaling & Bonner (1950)
500	0·178	—0·044	0·057	—0·045	—0·073	—0·003	
600	0·202	0·018	0·044	—0·030	—0·008	0·015	
	0·212	0·036	—0·004	—0·076	—0·027	0·013	Whaling & Bonner (1950)
700	0·224	0·024	0·038	—0·045	—0·058	0·002	
800	0·203	0·050	—0·005	—0·033	—0·014	—0·035	
900	0·178	0·059	0·057	—0·122	0·004	—0·011	
1000	0·275	0·148	0·013	—0·118	—0·037	—0·017	Whaling & Bonner (1950)

It is unfortunate that these results do not provide more useful data for theoretical calculations. If, however, the theoretical distributions of the two groups are arrived at separately, the present results should provide a more reliable check of the calculations than the results for the individual proton groups.

(ii) *Distribution of Long and Short Range Groups.*—Histograms are used to separate the total proton emission into long and short range groups as already described. The ratios of long to short range protons obtained are in general

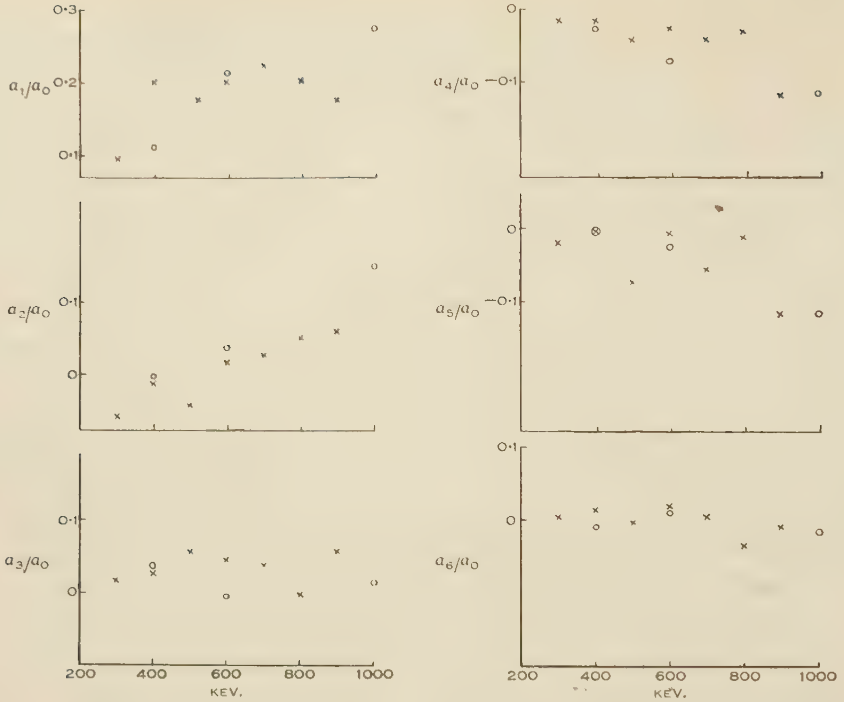


Fig. 5.—Variation with energy of the ratios a_r/a_0, $r=1, 2 \ldots 6$ for total proton emission from $Li^6(d,p)Li^{7*}$, Li^7.
○ Whaling and Bonner. × Authors.

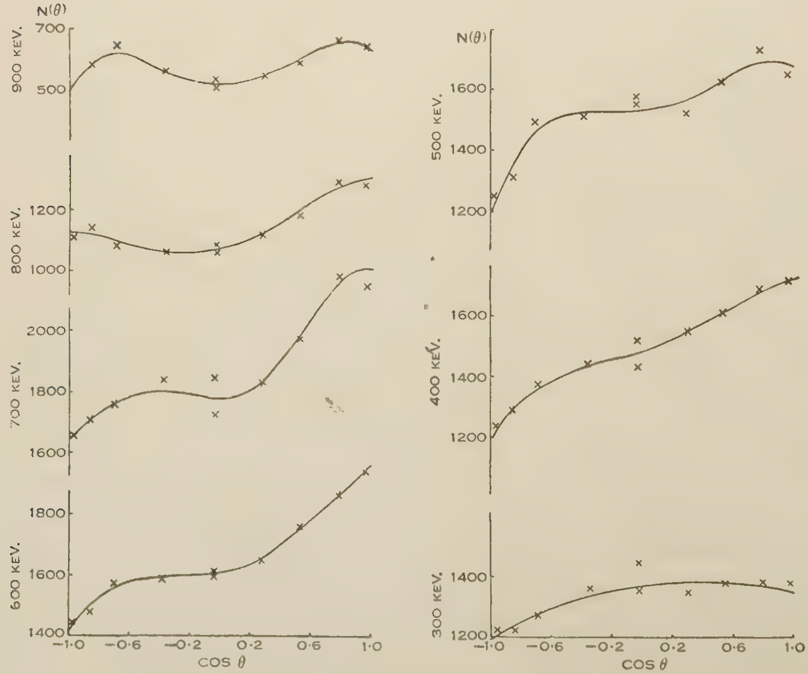


Fig. 6.—Angular distribution of long range protons from $Li^6(d,p)Li^7$.

agreement with those of Rumbaugh, Roberts, and Hafstad (1938) and Whaling and Bonner (1950) as far as may be judged from the graphical presentation of their results. The number of long range protons as a percentage of the total emission is shown in Table 2 for the various experiments. In the present experiment the separation of the two proton groups at 900 keV. and higher is complicated by the large number of spurious tracks due to the reaction Al$^{27}(d,p)$Al^{28}. Since the percentage of Li^6 proton tracks is never less than 60 per cent. of the total tracks counted, a correction to the 900 keV. results has been made using the observed emission from a blank aluminium-leaf target. This correction naturally reduces the accuracy of the 900 keV. results. At 1000 keV. the number of spurious tracks is so large that a reasonable correction is impossible. The supply of Li^6 became exhausted before a silver leaf backing to the target could be used. It is hoped, however, to repeat the high energy results when the Li^6 is again obtainable.

TABLE 2

PERCENTAGE OF LONG RANGE PROTONS FROM Li$^6(d,p)$Li^{7*}, Li^7

Deuteron Energy (keV.)	Rumbaugh, Roberts, and Hafstad (1938)	Whaling & Bonner (1950)	Krone, Hanna, & Inglis (1950)	Present Results
200	78	—	—	82
300	78	—	—	80
400	74	75	81	78
430	—	—	79	77
500	69	—	—	75
540	—	—	75	74
600	70	73	—	73
650	—	—	75	72
700	65	—	—	71
780	—	—	74	69
800	53	—	—	68
900	54	—	—	66
1000	—	71	—	63

The relative intensities of the long and short groups at different angles and energies are shown in Figures 6 and 7.

Expressions of the form

$$N(\theta)=a_0P_0+a_1P_1+a_2P_2+\ldots.+a_6P_6$$

have again been fitted to the points using numerical integration to evaluate the coefficients, a_n.

The results shown in Figures 8 and 9 indicate that a_5 and a_6 are essentially zero within the limits of experimental error. This is confirmed by other published data (Krone, Hanna, and Inglis 1950; Whaling and Bonner 1950). The curves in Figure 6, above 300 keV., have been drawn using the expression

$$N(\theta)=a_0P_0+a_1P_1+a_2P_2+a_3P_3+a_4P_4,$$

which is seen to provide a satisfactory fit to the points. The 300 keV. curve requires terms only to a_2P_2. The short range proton curves in Figure 7, however, are represented by expressions of the form

$$N(\theta)=a_0P_0+a_1P_1+a_2P_2.$$

These results are in contrast to those of Krone, Hanna, and Inglis (1950), who find a simpler distribution for the long range protons than for the short range group. The results presented here are in better agreement with those of Whaling and Bonner (1950) who obtain rather similar distributions for both long and short range groups.

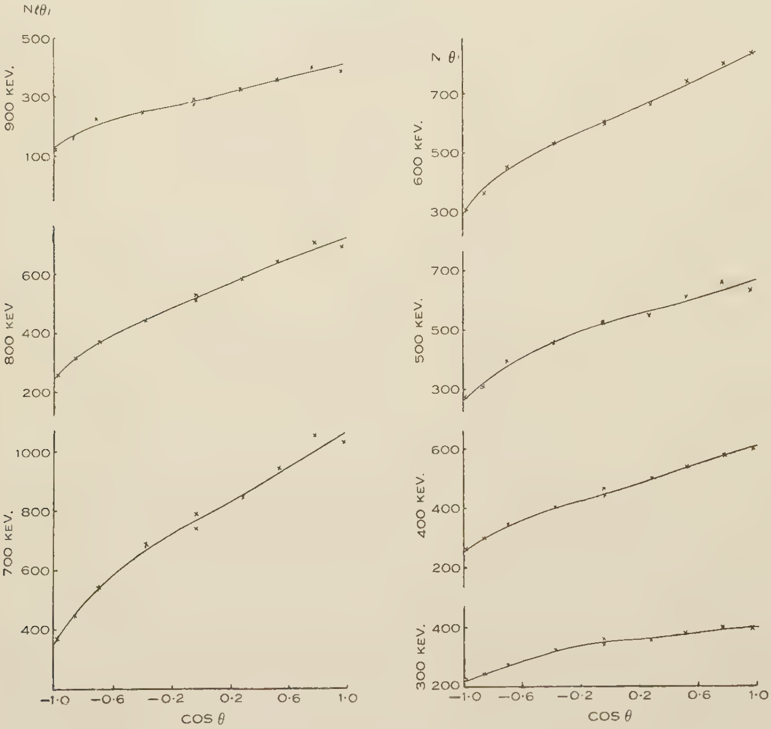


Fig. 7.—Angular distribution of short range protons from $Li^6(d,p)Li^{7*}$.

The present results have also been analysed* by fitting a polynomial of the sixth degree in cos θ to the experimental points by the method of least squares. Upon testing the significance of the various coefficients, A_n, of $\cos^n\theta$ for the long range protons it is found that only the coefficient, A_1, is significant, i.e. the curve of best fit is a straight line given by

$$N(\theta)=N(90^\circ)(1+A_1\cos\theta).$$

A similar analysis for the short range protons shows that in this case A_2 is also significant, i.e. the distribution may be expressed as

$$N(\theta)=N(90^\circ)(1+A_1\cos\theta+A_2\cos^2\theta).$$

---

* This analysis was carried out by the Department of Statistics, University of Melbourne.

However, in view of the general agreement in the shape of the long range proton curves of Figure 6 with those of Whaling and Bonner (1950), it is probable that the more complex expression involving powers of cos θ as high as the fourth is justified. The observed similarities in nearby curves, which are not taken into account in the least squares analysis, also support this view.

TABLE 3

VALUES OF a_1/a_0, a_2/a_0, a_3/a_0, . . . FOR LONG RANGE PROTONS

Energy (keV.)	a_1/a_0	a_2/a_0	a_3/a_0	a_4/a_0	a_5/a_0	a_6/a_0	Reference
	0·148	—0·014	0·054	—0·040	<0·010	—0·009	W. & B. (1950)
	0·047	—0·014	0·018	—0·035	0·046	—0·029	K., H., & I. (1950)
430	0·016	—0·004	0·021	—0·031	0·042	—0·028	K., H., & I. (1950)
500	0·113	—0·025	0·055	—0·054	—0·045	—0·016	
540	0·106	0·024	—0·051	—0·027	0·024	—0·009	K., H., & I. (1950)
600	0·127	0·041	0·041	—0·031	—0·007	0·017	
	0·138	0·096	—0·007	—0·046	—0·035	—0·018	W. & B. (1950)
650	0·105	0·147	0·068	0·008	—0·020	0·002	K., H., & I. (1950)
700	0·132	0·062	0·039	—0·051	—0·046	0·006	
780	0·205	0·146	—0·010	—0·150	—0·010	—0·012	K., H., & I. (1950)
800	0·091	0·098	—0·012	—0·029	—0·010	—0·027	
900	0·036	0·125	0·067	—0·147	0·036	—0·023	
1000	0·189	0·117	0·014	—0·044	—0·041	—0·056	W. & B. (1950)

TABLE 4

VALUES OF a_1/a_0, a_2/a_0, a_3/a_0, . . . FOR SHORT RANGE PROTONS

Energy (keV.)	a_1/a_0	a_2/a_0	a_3/a_0	a_4/a_0	a_5/a_0	a_6/a_0	Reference
300	0·246	—0·078	0·043	0·012	—0·062	—0·012	
400	0·376	—0·043	0·021	—0·011	—0·003	0·010	
	<0·010	0·026	—0·026	<0·010	—0·026	—0·018	W. & B. (1950)
	0·202	—0·152	—0·047	—0·141	—0·121	0·089	K., H., & I. (1950)
430	0·089	—0·240	—0·019	+0·204	0·010	0·081	K., H., & I. (1950)
500	0·373	—0·074	0·037	—0·050	—0·058	—0·015	
540	0·186	—0·193	—0·102	—0·077	0·005	0·067	K., H., & I. (1950)
600	0·416	—0·044	0·049	—0·033	—0·018	0·015	
	0·405	—0·122	<0·010	—0·156	—0·010	<0·010	W. & B. (1950)
650	0·140	0·061	—0·029	—0·241	—0·050	0·015	K., H., & I. (1950)
700	0·446	—0·060	0·036	—0·038	—0·075	—0·001	
780	0·428	0·089	0·063	—0·265	—0·046	—0·032	K., H., & I. (1950)
800	0·448	—0·059	—0·001	—0·038	—0·023	—0·036	
900	0·429	—0·041	0·056	—0·108	—0·017	—0·023	
1000	0·452	0·216	0·011	—0·266	—0·025	0·066	W. & B. (1950)

The variations in a_1/a_0, a_2/a_0, a_3/a_0, etc. with energy are shown in Table 3 and Figure 8 for the long range protons, and in Table 4 and Figure 9 for the short range group. The results of both Whaling and Bonner (1950) and Krone, Hanna, and Inglis (1950) have been included in the tables and figures. The quoted results of Krone, Hanna, and Inglis (1950) have been calculated by numerical integration from the data given in their Table 1. This calculation has been necessary in order to compare the results since they analyse their data in terms of the expression

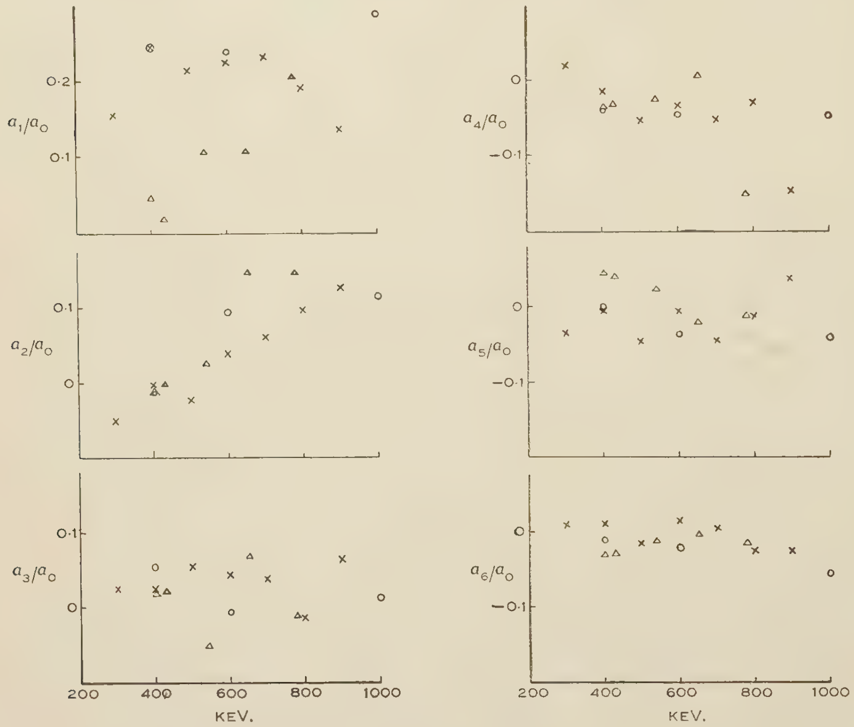


Fig. 8.—Variation with energy of the ratios A_r/A_0, $r=1, 2 \ldots 6$ for long range protons from $Li^6(d,p)Li^7$.

○ Whaling and Bonner. △ Krone, Hanna, and Inglis. × Authors.

$$N(\theta)=N(90^\circ)(1+A_1\cos\theta+A_2\cos^2\theta+\ldots+A_6\cos^6\theta).$$

A comparison of the results in Figures 8 and 9 reveals that all three experiments show reasonable agreement in the values of the ratios a_n/a_0 save in the following cases :

(*a*) a_1/a_0 for long range protons

The published results (Krone, Hanna, and Inglis 1950 ; Whaling and Bonner 1950) show a gradual increase with energy while the present results for a_1/a_0 show a falling off at 800 and 900 keV.

(*b*) a_4/a_0 for long range protons

In view of the close agreement of the other results the low values at 780 keV. (Krone, Hanna, and Inglis 1950) and 900 keV. (present results) are not regarded as being significant.

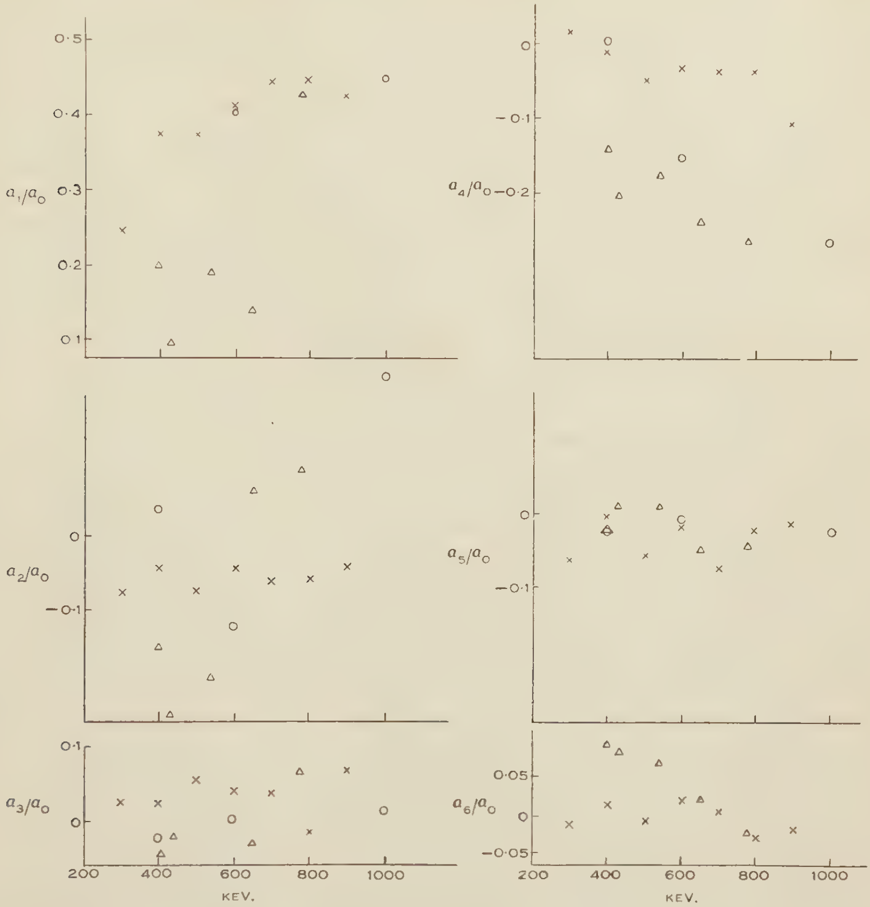


Fig. 9.—Variation with energy of the ratios, a_r/a_0, $r=1, 2....6$ for short range protons from $\mathrm{Li}^6(d,p)\mathrm{Li}^{7*}$.
○ Whaling and Bonner. △ Krone, Hanna, and Inglis. × Authors.

(*c*) a_1/a_0 for short range protons

Except for the very small value of Whaling and Bonner (1950) at 400 keV. (not shown in Fig. 9) the present results are in good agreement with theirs, whereas only the 780 keV. point of Krone, Hanna, and Inglis (1950) agrees with the other results.

(*d*) a_2/a_0 for short range protons

This ratio provides the greatest discrepancy in the three sets of results. Both the published experiments give rise to very erratic values, while the present results form a particularly consistent set. Since, moreover, they lie in the midst of the values of Whaling and Bonner (1950), and Krone, Hanna, and Inglis (1950) it is probable that they are a better representation of a_2/a_0.

(*e*) a_4/a_0 for short range protons

All three experiments reveal a gradual decrease in a_4/a_0 with increasing deuteron energy but no two agree on the actual values of the ratio. Since each of the three sets is internally consistent further measurements are needed to arrive at the best value of a_4/a_0.

From this comparison it appears that the differences observed in the shapes of the relative intensity distribution curves for the short range protons show up chiefly in the ratios a_1/a_0 and a_2/a_0. The agreement between Whaling and Bonner (1950) and the present results suggests that the angular distribution of short range protons is not more complex than the long range proton distribution. There is no evidence in either experiment for the approximate spherical symmetry in the long range protons observed by Krone, Hanna, and Inglis (1950).

If the differential cross section, $\sigma(\theta)$, is expressed in the form

$$\sigma(\theta)=\Sigma a_n P_n\ (\cos\theta),$$

where P_n (cos θ)=Legendre polynomial of order n,

and a_n=coefficient of nth order polynomial,

it has been shown by Wolfenstein (1949) that all the coefficients a_n with $n>2l_{max}$, are zero. In this case l_{max} is the maximum orbital momentum of the incident deuteron. The long range proton distribution above 300 keV. may thus be accounted for by *s*, *p*, and *d* deuterons while only *s* and *p* deuterons are required to describe the short range proton results. The need to invoke the use of *d* deuterons at these low energies is in agreement with the observed results in the $Li^6(d,\alpha)He^4$ reaction. With incident deuterons of $l=0$ or 1 the compound nucleus may have an angular momentum, $j=0$, 1, 2, or 3 while $l=2$ deuterons introduce only the additional state of $j=4$.

The departure from symmetry about the plane through the target normal to the incident beam is due to the occurrence of odd powers of cos θ in the expression for $\sigma(\theta)$, which is given by the dispersion formula

$$\sigma_{PQ}(\theta) \sim (\lambda/2\pi)^2 \left| \frac{\sum\limits_{R} H_{PR} H_{RQ}}{(E-E_R)+i/2\,.\,\Gamma_R} \right|^2.$$

The matrix elements $H_{PR}H_{RQ}$ may be expressed in terms of Legendre polynomials so that

$$\sigma_{PQ}(\theta) \propto |\Sigma b_r P_r|^2.$$

The odd powers of cos θ may only arise from products of the type P_{l_1} (cos θ) . P_{l_2} (cos θ) when either l_1 or l_2 (but not both) is odd. This means that states of Be^8 of different parities are concerned in the reaction.

Although the simpler distribution of the short range protons suggests a spin of $\frac{1}{2}$ for the excited state of Li^{7*}, confirmation of the experimental results seems desirable before a complete theoretical interpretation is attempted. The present results do not support the suggestion of Hanna and Inglis (1949) that the Li^{7*} state has a spin of 5/2.

ANGULAR DISTRIBUTION OF PARTICLES FROM $Li^6 + D$ REACTION

Typical field of view in microscope, showing α-particles and indistinct proton tracks.

IV. ACKNOWLEDGMENTS

The authors wish to thank Professor L. H. Martin for his advice and encouragement in the course of this work ; and Dr. J. C. Bower for helpful discussions. The photographic plate measurements have been carried out by Miss E. McKenzie with the assistance of Miss M. Stringer and Mrs. D. Greig. The Li^6 targets were prepared by Mr. J. R. Bird. The continued financial assistance of the Commonwealth Scientific and Industrial Research Organization and Dunlop Rubber Australia Ltd. is gratefully acknowledged.

V. REFERENCES

GUEST, P. G. (1950).—*Phil. Mag.* **41** : 124.

HANNA, S. S., and INGLIS, D. R. (1949).—*Phys. Rev.* **75** : 1767.

HEYDENBURG, N. P., HUDSON, C. M., INGLIS, D. R., and WHITEHEAD, W. D. (1948).—*Phys. Rev.* **74** : 405.

KRONE, R. W., HANNA, S. S., and INGLIS, D. R. (1950).—*Phys. Rev.* **80** : 603.

MARTIN, L. H., BOWER, J. C., DUNBAR, D. N. F., and HIRST, F. (1949).—*Aust. J. Sci. Res.* A **2** : 25.

RESNICK, R., and INGLIS, D. R. (1949).—*Phys. Rev.* **76** : 1318.

RUMBAUGH, L. H., ROBERTS, R. B., and HAFSTAD, L. R. (1938).—*Phys. Rev.* **54** : 657.

WHALING, W., and BONNER, T. W. (1950).—*Phys. Rev.* **79** : 258.

WOLFENSTEIN, L. (1949).—*Phys. Rev.* **75** : 1664.

FURTHER MEASUREMENTS OF THE ANGULAR DISTRIBUTION OF α-PARTICLES FROM THE $Li^7(p, \alpha)He^4$ REACTION

By F. Hirst* and R. Uebergang*

[*Manuscript received March* 30, 1951]

Summary

The angular distribution of the α-particles from the $Li^7(p,\alpha)He^4$ reaction has been measured for a proton energy range 0·3 to 0·9 MeV. using scintillation counter technique. A statistical analysis of the observations has been undertaken and results have been compared with those of other experiments. The angular distribution is well represented by the expression

$$Y(\theta)=Y(90^\circ)[1+A(E)\cos^2\theta+B(E)\cos^4\theta],$$

where $A(E)$ and $B(E)$ are energy dependent parameters denoting the departure from spherical symmetry and implying that p and f protons are important in the reaction. Curves of $A(E)$ and $B(E)$, of the theoretical form given by Inglis, have been fitted to the combined data of recent experiments analysed according to the above expression.

I. Introduction

Results of early experiments (Young, Ellett, and Plain 1940 ; Swartz *et al.* 1944 ; Rubin, Fowler, and Lauritsen 1947), which detected the departure from spherical symmetry of the angular distribution of the α-particles from the $Li^7(p,\alpha)He^4$ reaction, differed considerably as to the dependence of the asymmetry on energy. Also, to explain this asymmetry, it was assumed that only p protons took part in the reaction, in which case the angular distribution could be expressed by the relation

$$Y(\theta)=Y(90^\circ)[1+A_0(E)\cos^2\theta].$$

Here $Y(\theta)$ represents the yield per unit solid angle at an angle θ to the direction of the incident proton beam, both $Y(\theta)$ and θ being measured in centre of mass coordinates ; E is the proton energy in laboratory coordinates and $A_0(E)$ is an energy dependent parameter which is a measure of the asymmetry of the angular distribution.

Later experiments of Heydenberg *et al.* (1948*a*, 1948*b*), using proton energies above 1 MeV., indicated that the angular distribution could be represented more precisely by the expression

$$Y(\theta)=Y(90^\circ)[1+A(E)\cos^2\theta+\mathrm{B}(E)\cos^4\theta].$$

$B(E)$ is a second energy dependent parameter to account for f protons reacting at higher bombarding energies. It was shown by Martin *et al.* (1949) and confirmed by Talbott, Busala, and Weiffenbach (1950*b*) that this term remained significant down to energies of the order of 0·5 MeV.

---

* Physics Department, University of Melbourne.

The angular distribution has been considered theoretically by Critchfield and Teller (1941) who have developed the theory for p protons and this has been extended by Inglis (1948) to include f protons. The theoretical treatment was based on the Breit-Wigner dispersion formula on the assumption that the observed asymmetry arose from interference of α-particle waves arising from two levels of the Be^8 compound nucleus, a broad level with angular momentum $J=0$ and a sharp level with angular momentum $J=2$. The fact that the observed distribution is most readily explained by assuming p and f protons enter the reaction is evidence in support of the assignment of odd parity to the ground state of the Li^7 nucleus.

In the investigations mentioned previously, counting of the α-particles was performed by ionization chamber (Young, Ellett, and Plain 1940; Heydenberg *et al.* 1948*a*), proportional counter (Heydenberg *et al.* 1948*b*), and nuclear emulsion photographic plates (Swartz *et al.* 1944; Rubin, Fowler, and Lauritsen 1947; Martin *et al.* 1949; Talbott, Busala, and Weiffenbach 1950*b*). In the present experiment scintillation counter technique has been used. Ease of counting was a feature of this method, whilst the counting equipment could be adjusted to give a negligible background count.

II. METHOD

(*a*) *Electrostatic Generator*

Energetic proton beams were obtained by means of a 1 MV. electrostatic generator described previously (Martin *et al.* 1949). Prior to the commencement of the present experiment, this generator was equipped with voltage stabilization and current regulation of the magnetic spectrograph.

Voltage stabilization was achieved by variation of the corona current down the outside of the acceleration tube, this corona current being controlled by automatic adjustment of the potential on insulated discharge points situated at the bottom of the tube column. The error signal for the voltage stabilizer was derived from two parallel insulated metal jaws bounding the horizontal exit slit of the magnetic analyser. When the high voltage electrode potential altered, the proton beam struck either the upper or the lower jaw, depending on whether the beam energy suffered a decrease or an increase respectively. The signal from the jaws controlled the output voltage of an R.F. power supply (5–12 kV. negative) which was applied to the insulated discharge points controlling the corona loading. By this method of stabilization, fluctuations up to ±10 kV. can be eliminated.

The magnetic analyser regulator circuit followed one designed by Caro and Parry (1949) for use in a cosmic ray spectrograph. A single feed-back loop is not satisfactory since the inductive time constant of the magnet coil makes it difficult to obtain good performance without " hunting ". Thus the regulator has a fast feed-back loop regulating the voltage across the coil by controlling the field current of the D.C. generator which is fed from a separate supply. A second slow acting loop then corrects for current changes by adjusting the value at which the fast loop operates. The signal for the current loop is derived from the potential difference across tungsten filament light globes (in series

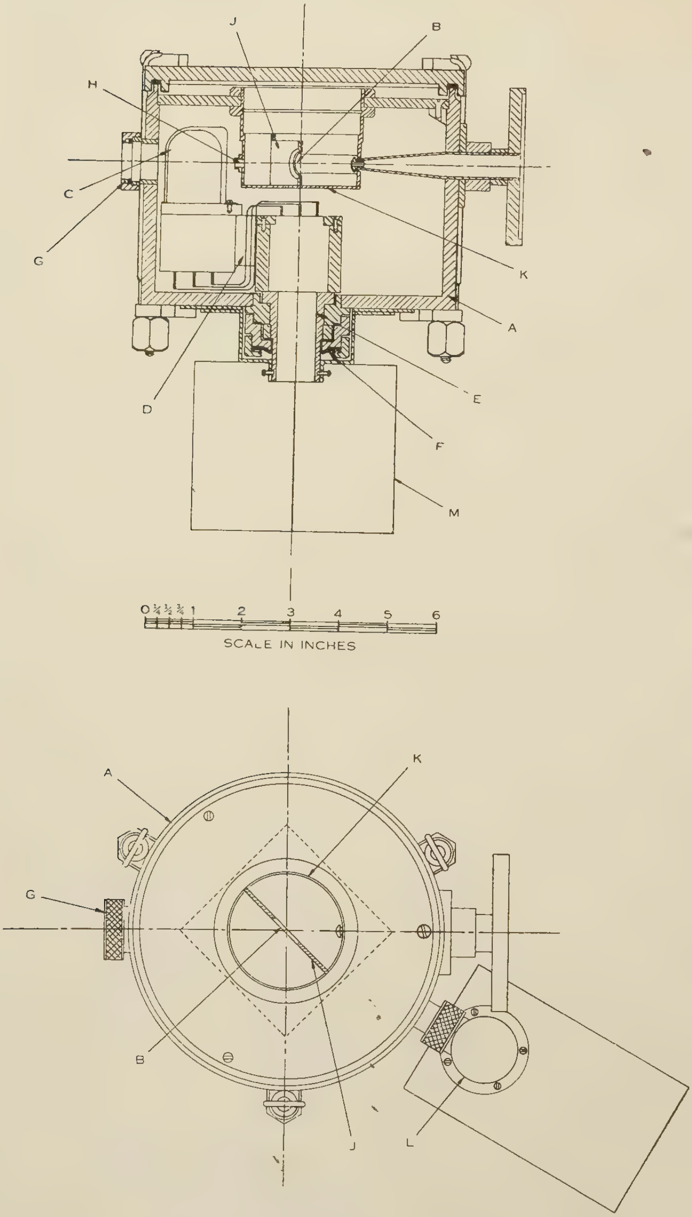

J
B
H
C
G
K
A
E
D
F
M
0 ¼ ½ ¾ 1 2 3 4 5 6
SCALE IN INCHES
K
A
G
B
J
L

Fig. 1.—Counting head.

A, Vacuum chamber ; *B*, Target ; *C*, Search counter ; *D*, Arm ; *E*, Spindle ; *F*, Wilson seal ; *G*, Inspection window ; *H*, Quartz disc ; *J*, Target frame ; *K*, Target holder ; *L*, Monitor ; *M*, Search amplifier.

with the magnet current), which are insensitive to ambient temperature since most of their dissipation is by radiation (Minett, MacRae, and Townsend 1949). In normal operation the short term overall wander of the regulated current is 0·1 per cent., thereby limiting the variation of the proton energy incident on the target to less than 2 keV.

(*b*) *Scintillation Counter*

The counting head used in the experiment is shown in Figure 1. It consisted of a cylindrical vacuum chamber (A), containing a target (B) mounted axially, about which a search counter (C) rotated in the plane containing the proton beam. The search counter was mounted on an arm (D) attached to a spindle (E), which was brought through the chamber wall by means of a Wilson seal (F). A protractor mounted on the spindle outside the vacuum chamber recorded the angular position of the search counter, α-particles being counted in the angular ranges (13–90°, 90–164°).

Accurate machining of the counting chamber was essential to make certain that the centre of revolution of the counting area was coincident with the target spot, any departure from coincidence being less than 0·1 mm. The counting chamber was connected to the exit port of the magnetic spectrograph by a sylphon bellows to allow the chamber to be aligned with the proton beam. This adjustment was accomplished by observing, through an inspection window (G) in the vacuum housing, that the proton beam was centred on a small quartz disc (H). The counting head was then clamped in position, and the quartz disc and inspection window covered to prevent any extraneous light falling on to the scintillation screen.

The screen of the search counter was prepared by depositing zinc sulphide (silver activated) phosphor on the glass envelope of a 931A photoelectron multiplier, by sedimentation of the phosphor suspended in water. A weighed amount of zinc sulphide was used to obtain an optimum screen thickness, sodium silicate being added as a binder together with a mild electrolyte, ammonium carbonate, which ensured that the particles settled uniformly.

The search photoelectron multiplier was placed within a metal housing which acted as an electrostatic shield. An aperture (7·5 by 5 mm.) in the metal housing allowed α-particles to impinge on the scintillation screen adjacent to the area of maximum sensitivity of the photo-cathode, this aperture being covered with aluminium leaf to exclude light. The mean distance from the aperture to the target spot was 4 cm.

During preliminary observations, appreciably fewer counts were recorded at backward angles than were expected from the forward and backward symmetry of the distribution. Controlled experiments, using a polonium α-particle source, showed that this anomalous counting was caused by the leakage flux of the magnetic spectrograph overriding the electron optics of the 931A photoelectron multiplier. To overcome this effect the counting head was moved further away from the magnet and a mu-metal shield was placed around the photoelectron multiplier. With these modifications the operation of the search counter was shown to be independent of angular position within the limits of statistical variation.

The α-particle flux (a function of the beam current and the target thickness) was monitored by another scintillation counter (L, Fig. 1) set at a fixed angle. In the case of the monitor the phosphor screen was deposited on a glass disc, which was cemented by canada balsam to the polished end of a short optical glass light pipe, the cylindrical walls of which had been silvered and blackened.

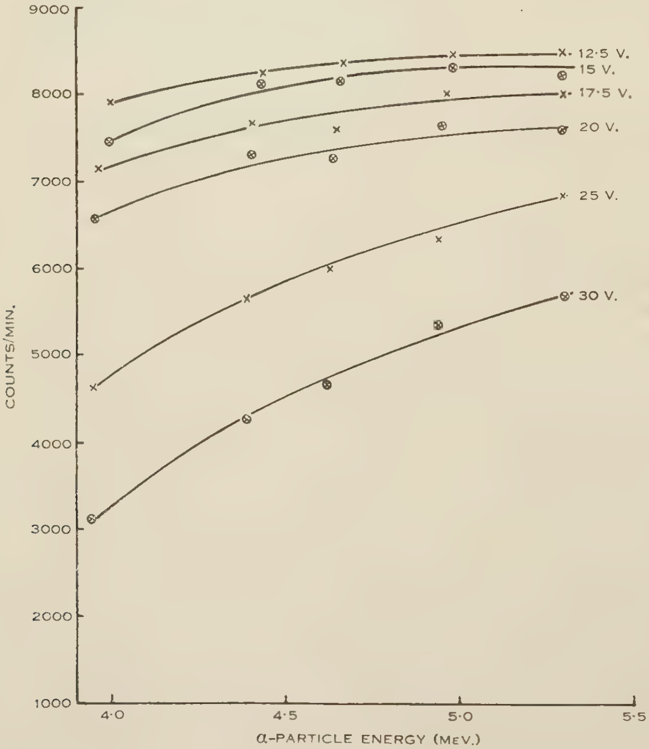


Fig. 2.—Counting rate against α-particle energy.

Scintillations from this screen were guided by the light pipe to a 27 M1 photo-electron multiplier situated outside the vacuum chamber. Here again the scintillation screen was covered with aluminium leaf to exclude light.

(c) *Independence of Counting Rate on α-Particle Energy*

In the development of the scintillation counter for the present angular distribution measurements, an important consideration which required investigation was the dependence of counting efficiency on α-particle energy, since the energy of the emitted α-particles decreases continually from a maximum at 0° to a minimum at 180°. Here counting efficiency refers to the ratio

$$\frac{\text{Number of } \alpha\text{-particles counted}}{\text{Number of incident } \alpha\text{-particles}}.$$

Preliminary work with a polonium α-particle source using air as an absorber had shown that, for a discriminator bias of less than 15 volts, this efficiency became independent of α-particle energy above 4·5 MeV. Figure 2 shows the variation of counting efficiency with α-particle energy in the range 4·0 to 5·3 MeV. for different values of discriminator bias. Since the energy of the α-particles to be counted in this experiment was in the range 5·0 to 8·0 MeV., it appeared likely that the counting efficiency throughout the investigation would be independent of energy. Within the latter range, the independence of counting efficiency on energy was checked by observing the counting rate at a fixed angle (viz. 14°) after α-particles from the reaction had traversed respectively one and two thicknesses of cellophane absorber (3 cm. air equivalent). During this test the proton beam was maintained at 0·5 MeV., giving at the fixed angle an initial α-particle energy of 9·3 MeV. which was reduced to 7·4 MeV. on traversing one thickness of absorbing screen, and was further reduced to 5·0 MeV. after penetrating two cellophane screens. It was found that the 5·0 MeV. and 7·4 MeV. α-particles were counted with equal efficiency and it seemed safe to conclude that, for the α-particle energies encountered in the present investigation, counting efficiency could be considered independent of angular position.

(*d*) *Electronics*

The signals from the photoelectron multipliers were fed to separate head amplifiers mounted directly on the counting chamber, these amplifiers being supplied by independent stabilized high tension units (300 V.). The search amplifier (*M*, Fig. 1) was mounted on a bracket attached to the spindle which carried the search counter and rotated with it. The amplified signals were then led to separate discriminators and scalers. A bank of Minimax dry batteries supplied the high tension (up to 1000 volts) for the voltage dividers of the photoelectron multipliers. In the case of the search counter, the connections to the photoelectron multiplier electrodes were brought outside the vacuum chamber through Kovar seals set in the end of the spindle. To overcome an excessive background count caused by discharge inside the vacuum chamber from the high voltage connections of the photoelectron multiplier, its leads and pins were painted with glyptal. This procedure eliminated the background count when the pressure in the vacuum chamber was $\simeq 10^{-5}$ mm. Hg.

(*e*) *Target Details and Counting Technique*

Thin lithium targets (<20 keV. thick) were made by evaporating lithium hydroxide heated on a platinum foil on to thin aluminium leaf cemented to an aluminium frame. The thickness of a typical target was determined from the half width of the 440 keV. γ-ray resonance and was found to be 17 keV. To locate the target at its fixed position, the target frame (*J*, Fig. 1) was mounted in a holder (*K*, Fig. 1) which was rigidly attached to the vacuum chamber walls, the target frame being easily removed for target renewal. Since the target frame was mounted at 45° to the incident proton beam and the beam was limited by a 2 mm. diameter circular aperture, the target spot was an ellipse with axes

2·8 and 2 mm. Windows covered with red cellophane were provided in the target holder to prevent scattered protons from striking the scintillation screens. In the angular range 13 to 90°, the α-particles also passed through the aluminium foil (2 mm. air equivalent) forming the target backing.

At each bombarding energy eight runs were performed, a new target being used for each run. A run consisted of recording the α-particle count at each of ten selected angles against a monitor count of 300. For successive runs, angles were counted in opposite order so tending to balance out any effect arising from altering target condition.

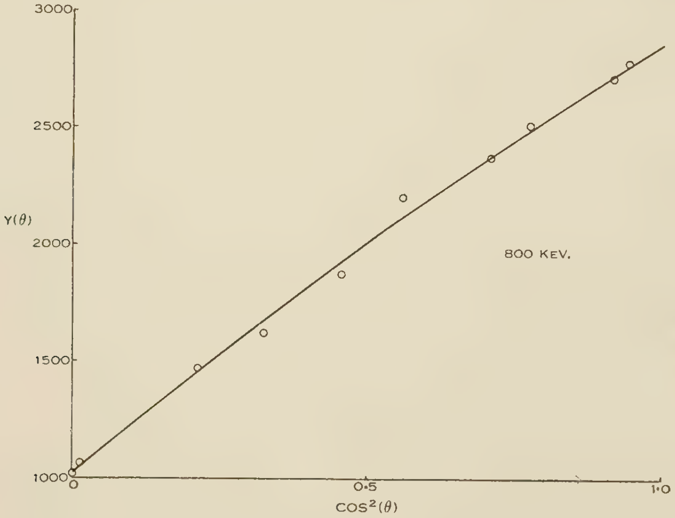


Fig. 3.—$Y(\theta)$ against $\cos^2(\theta)$.

III. Results

For each bombarding energy the eight counts at each angle were added and normalized to a uniform monitor count. These angles measured in laboratory coordinates were converted to centre of mass coordinates. Also, the total count at each angle measured for the same solid angle in laboratory coordinates

Table 1

E (keV.)	300	350	400	450	500	550	600 (i)	600 (ii)	700	800	900
$A(E)$ ..	0·44	0·49	0·80	1·01	1·04	1·37	1·46	1·63	1·83	2·02	2·11
Probable error $A(E)$ ..	0·12	0·17	0·09	0·10	0·09	0·17	0·20	0·21	0·29	0·15	0·22
$B(E)$ ..	—0·04	—0·16	—0·05	—0·10	—0·01	—0·07	—0·13	—0·27	—0·27	—0·22	—0·28

was converted to a standard solid angle in the centre of mass system. Formulae for these two conversions are given by Martin *et al.* (1949).

The parameters $A(E)$ and $B(E)$ involved in the expression for the angular distribution were evaluated from the recorded observations for a given particle energy by the method of least squares using orthogonal polynomials after the procedure of Guest (1950). Figure 3 shows a curve fitted by the least squares method of Guest to the $Y(\theta)$ recordings obtained for a typical proton energy (viz. 800 keV.) plotted against $\cos^2\theta$. The values of $A(E)$ and $B(E)$ obtained are tabulated in Table 1 and are plotted against E in Figure 4.

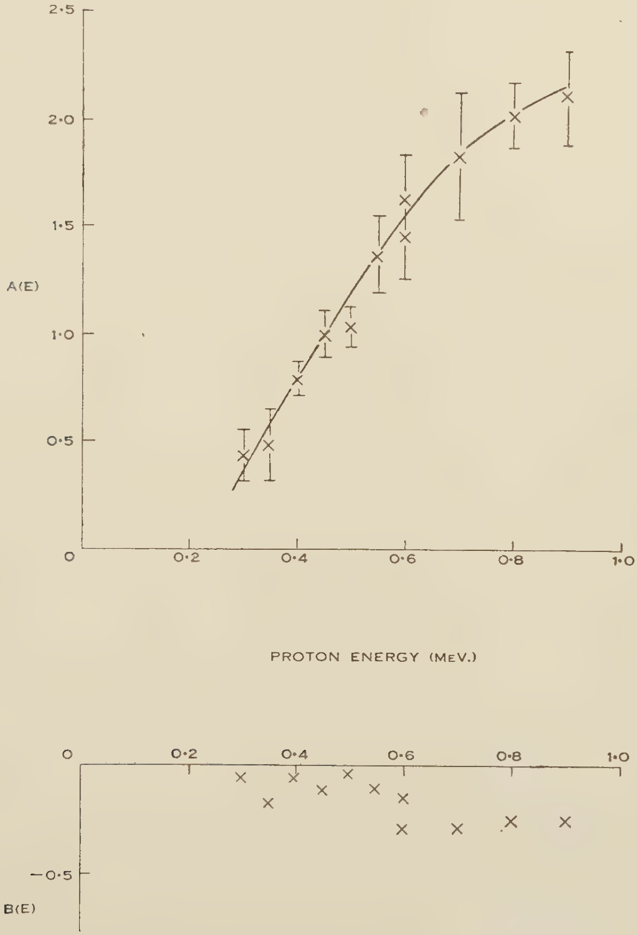


Fig. 4.—$A(E)$ and $B(E)$ against proton energy.

Detailed observations of the photographic plate data from the experiment by Martin *et al.* (1949) were available, and curves were also fitted to these measurements again by the method of Guest (1950), the values of $A(E)$ and $B(E)$ obtained being given in Table 2.

Table 2

E (keV.)	140	275	370	440	520	625	715	810	870	960
$A(E)$ ..	—0·44	0·54	1·05	1·05	1·14	1·90	2·24	2·32	2·69	2·82
$B(E)$ ..	+0·61	—0·11	—0·37	—0·30	—0·07	—0·44	—0·72	—0·51	—0·64	—0·46

The results of the scintillation counter experiment (Table 1) are shown in Figure 5 combined with those of Martin *et al.* (Table 2) and of other workers (Heydenberg *et al.* 1948*a*, 1948*b* ; Talbott, Busala, and Weiffenbach 1950*b*).

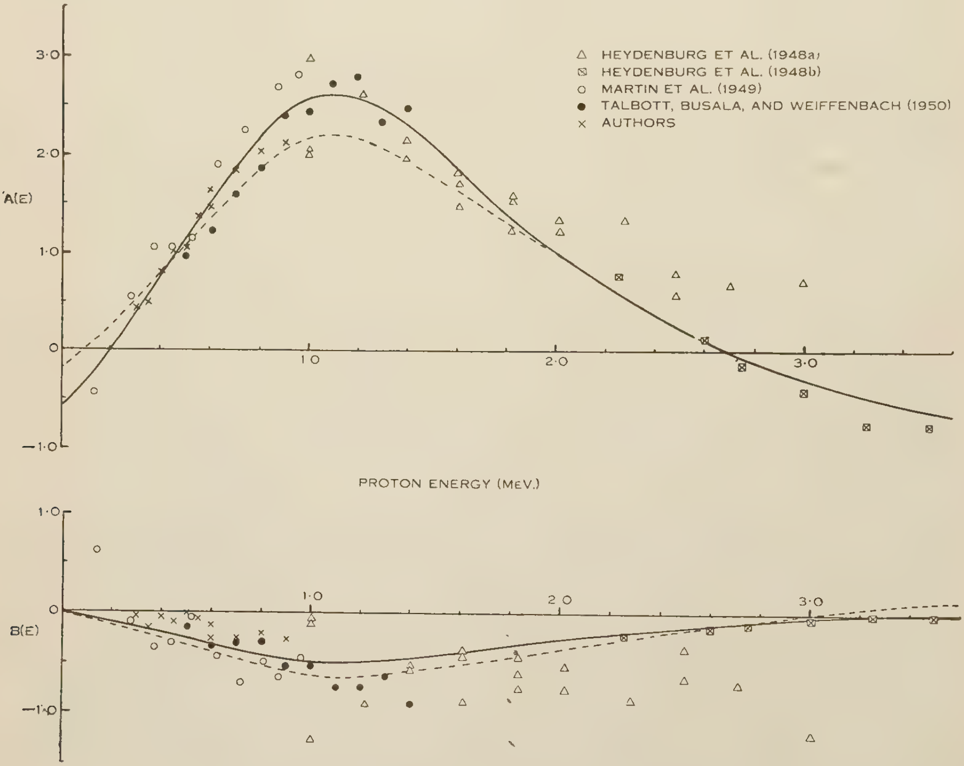


Fig. 5.—$A(E)$ and $B(E)$ against proton energy.

IV. Discussion

The present investigation was undertaken to obtain more information on certain features observed in the results of earlier workers. The experimental work of Heydenberg *et al.* (1948*a*, 1948*b*) appears to show two trends in the values of $A(E)$ near 1 MeV., these authors favouring the lower $A(E)$ values, whilst the results of Martin *et al.* (1949) are in closer agreement with the higher values. As may be seen from Figure 5, our values of $A(E)$ for the higher proton energies are lower than those of Martin *et al.*, but the difference generally is within the expected statistical variation, thus emphasizing the inherent difficulty in fixing $A(E)$ precisely for a certain energy, unless an extremely large number of α-particle counts is made. Again the various sets of observed data in general fall into two divisions, either below 1 MeV. or above 1 MeV., and only in the experiment of Talbott, Busala, and Weiffenbach (1950*b*) does a set of data completely cover the range in which the maximum value of $A(E)$ lies. It is obvious that, for a systematic study of the $A(E)$ curve in this region, a generator which is capable of operating within the energy limits $0 \cdot 5$ to $1 \cdot 5$ MeV. is desirable. For this reason the consistent set of data presented by Talbott, Busala, and Weiffenbach must be given greater weight when considering the most likely variation of $A(E)$ with energy near 1 MeV.

In the energy region near $0 \cdot 5$ MeV., where the results of different observers are in close agreement, the $A(E)$ curve may be located with precision. Our results agree well with those of Talbott, Busala, and Weiffenbach especially when a correction for target thickness is applied to their results. These authors used 100 keV. thick targets and considered a reasonable correction for target thickness would be to move their $A(E)$ values 50 keV. along the abscissa in the direction of lower energy.

Our results suggest that the $A(E)$ curve cuts the abscissa at approximately $0 \cdot 2$ MeV. The energy where $A(E)$ is zero is of particular interest, since it imposes a severe limitation on the possible shape of the theoretical curve which can be fitted to the observed data. The measurements of $A(E)$ at very low energies are inadequate because compilation of data is rendered difficult by the poor yield of the reaction in this energy region. Further work has been undertaken with new equipment giving larger beam currents, to obtain more precise values for $A(E)$ at energies near $0 \cdot 2$ MeV.

Another aim of the present experiment was to investigate the reality of a trend noticed in the results of Young, Ellett, and Plain (1940), Martin *et al.* (1949) and preliminary data of Talbott, Busala, and Weiffenbach (1950*a*) suggesting the possibility of an inflexion in the $A(E)$ curve at approximately 450 keV. Our results show that if such a trend is real it is probably much smaller than the observations of the earlier workers have indicated, any variation of the $A(E)$ curve in this region being within the limits of statistical uncertainty.

The results of the present experiment also confirm the existence of a $B(E)$ term down to energies of the order of $0 \cdot 5$ MeV., the general trend of $B(E)$ values being similar to that observed by Martin *et al.*, and Talbott, Busala, and Weiffenbach. Since the $B(E)$ term represents the curvature of the $Y(\theta)/\cos^2\theta$ curve, its value is extremely sensitive to statistical variation in $Y(\theta)$. Although the data shown in Figure 5 cannot be expected to give the energy dependence of $B(E)$ with precision, they certainly support the assumption that f protons are reacting at energies well below 1 MeV.

V. THEORETICAL ANALYSIS

Inglis (1948), assuming that p and f protons are involved in the $Li^7(p,\alpha)He^4$ reaction, has derived a theoretical expression containing five disposable constants for the variation of $A(E)$ with energy, this expression having the form

$$A(E)=\frac{C_0E^2+C_1E+C_2}{E^2+C_3E+C_4}.$$

The broken line shown in Figure 5 is a curve satisfying the theoretical expression of Inglis and fitted by Heydenberg *et al.* (1948*b*) to the observed data then available. Above 1 MeV. this curve was fitted by the latter authors to their own data which were analysed according to the relation

$$Y(\theta)=Y(90^\circ)[1+A(E)\cos^2\theta+B(E)\cos^4\theta].$$

Below 1 MeV. the fitting was based on an arbitrary extrapolation of $A(E)$ values guided by the trend of the $A_0(E)$ curve. At that time the data below 1 MeV. lacked precision and had failed to detect the $B(E)$ term.

A new $A(E)$ curve of the same theoretical form has been fitted by the present authors to the results of Heydenburg *et al.* combined with the more recent data which have detected the existence of the $B(E)$ term, below 1 MeV. This curve (full line) is also shown in Figure 5. Restrictions imposed in fitting our theoretical curve were:

(i) $A(E)$ became zero at 0·2 MeV. and 2·7 MeV.
(ii) The maximum value of $A(E)$ was 2·6 at 1·1 MeV.
(iii) At 0·5 MeV., where the cluster of points allows $A(E)$ to be determined accurately, $A(E)$ was 1·1.

It is seen from Figure 5 that this theoretical curve has been fitted to agree most closely with the results of the present experiment below 0·8 MeV., with those of Talbott, Busala, and Weiffenbach (1950*b*) in the region of 1 MeV. and with the high energy values of $A(E)$ obtained by Heydenburg *et al.* near 2·7 MeV. This curve indicates in our opinion the best $A(E)$ curve, satisfying the theoretical expression of Inglis, that can be fitted to the observed data now available. The five disposable constants used in fitting this theoretical curve together with those used by Heydenburg *et al.* are listed in Table 3.

TABLE 3

	C_0	C_1	C_2	C_3	C_4
Authors	—1·74	5·03	—0·94	—1·73	1·66
Heydenburg *et al.* ..	—1·75	4·90	—0·47	—1·91	2·15

Inglis (1948) has derived also a theoretical expression for $B(E)$, which involves two further arbitrary parameters and has the form

$$B(E)=\frac{C_5E^2+C_6E}{E^2+C_3E+C_4}.$$

Figure 5 shows a $B(E)$ curve (broken line) conforming to this relation and fitted by Heydenburg *et al.* to the earlier data, together with our curve (full line) again satisfying the same expression fitted to the more recent data. Because of the very large probable error in the values of $B(E)$ and their consequent scatter, it is not possible to fit a curve to the observed $B(E)$ data accurately. The values of the additional parameters used in fitting the theoretical $B(E)$ curves are listed in Table 4.

TABLE 4

	C_5	C_6
Authors	0·20	—0·65
Heydenburg *et al.*	0·4	—1·2

VI. ACKNOWLEDGMENTS

The authors wish to express their thanks to Professor L. H. Martin for continued interest and advice throughout the course of this experiment. We also thank Dr. J. C. Bower and Mr. D. N. F. Dunbar for helpful discussion, and Mr. J. G. Campbell for building the current regulator for the magnetic analyser. We are indebted to Mr. C. Moncrieff for the extreme care exercised in the mechanical construction of the counting chamber. Continued financial assistance was provided by the Commonwealth Scientific and Industrial Research Organization.

VII. REFERENCES

CARO, D. E., and PARRY, J. K. (1949).—*J. Sci. Instrum.* **26** : 374.

CRITCHFIELD, C. L., and TELLER, E. (1941).—*Phys. Rev.* **60** : 10.

GUEST, P. (1950).—*Phil. Mag.* **41** : 124.

HEYDENBURG, N. P., HUDSON, C. M., INGLIS, D. R., and WHITEHEAD, W. D. (1948*a*).—*Phys. Rev.* **73** : 241.

HEYDENBURG, N. P., HUDSON, C. M., INGLIS, D. R., and WHITEHEAD, W. D. (1948*b*).—*Phys. Rev.* **74** : 405.

INGLIS, D. R. (1948).—*Phys. Rev.* **74** : 21.

MARTIN, L. H., BOWER, J. C., DUNBAR, D. N. F., and HIRST, F. (1949).—*Aust. J. Sci. Res.* A **2** : 25 : *Nature* **164** : 310.

MINETT, E. E., MACRAE, D. A., and TOWNSEND, J. (1949).—*Rev. Sci. Instrum.* **20** : 136.

RUBIN, S., FOWLER, W. A., and LAURITSEN, C. C. (1947).—*Phys. Rev.* **71** : 212.

SWARTZ, C. D., ROSSI, H. H., JENNINGS, B., and INGLIS, D. R. (1944).—*Phys. Rev.* **65** : 80.

TALBOTT, F. L., BUSALA, A., and WEIFFENBACH, G. C. (1950*a*).—Dep. Phys. Cath. Univ. Amer. Semi-Ann. Rep. (Feb.).

TALBOTT, F. L., BUSALA. A., and WEIFFENBACH, G. C. (1950*b*).—Dep. Phys. Cath. Univ. Amer. Tech. Rep. No. 1.

YOUNG, V. J., ELLETT, A., and PLAIN, G. J. (1940).—*Phys. Rev.* **58** : 498.

THE SURFACE LOSS CORRECTION FOR NUCLEAR EMULSIONS

By S. L. Martin*

[*Manuscript received February* 8, 1951]

Summary

When nuclear emulsions are used in the investigation of the angular distributions of scattered particles those tracks which pass through the emulsion surface cannot be accepted for measurement and in consequence the observed angular distribution must be corrected to allow for this. In the present paper, formulae are developed for the case of particles recoiling from a mono-energetic beam of particles incident at any angle on the surface. The effect of small-angle scattering is neglected but on the basis of work done by Powell and Occhialini it is believed that the formulae are valid for recoil energies at least as great as 5 MeV. Curves are given from which the factor can be obtained in any given case.

I. Introduction

In using nuclear emulsions for the determination of the angular distribution of scattered particles no tracks which pass through the surface can be accepted for measurement since their lengths are uncertain. This means that a correction factor, which for a given set of conditions is a function of the angle of scattering, must be applied to correct for the discrimination introduced because of the rejected tracks. Richards (1941) has derived an expression for the factor when a mono-energetic beam of incident particles is parallel to the emulsion surface and the scattering is isotropic in a centre-of-mass system of coordinates. The latter assumption which appears to be valid (Dee and Gilbert 1937 ; Barschall and Kanner 1940 ; Martin *et al.* 1950) in the case of *n–p* scattering, the one observed in Richards's experiments, limits the applicability of the expressions obtained. Champion and Powell (1945) have given a graphical method of making the correction when the incident beam is parallel to the surface. In the present paper the general case is considered and expressions are obtained for any type of scattering when the bombarding beam is incident at any angle on the emulsion. When these formulae are applied to the special case of the beam parallel to the surface it is found that simple formulae can replace the lengthy graphical procedure due to Champion and Powell.

II. Surface Loss Correction Factor

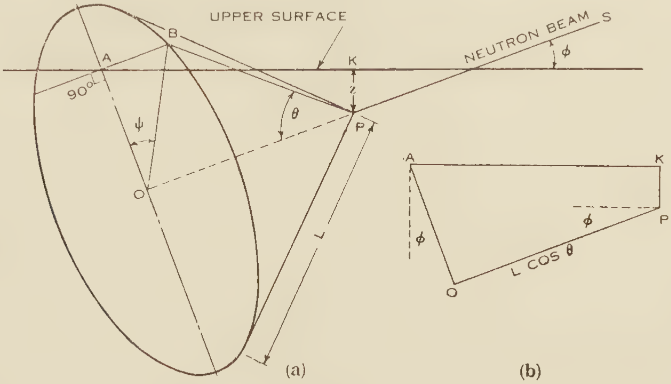
(*a*) *Loss Through the Upper Surface*

Consider a beam of mono-energetic particles entering the emulsion along a direction *SP* (Fig. 1 (*a*)) making an angle φ with the surface and suppose that a collision occurs at a point *P*, at a depth z below the surface. If we assume that the emulsion does not polarize either the incident beam or the recoils, it

---

* Department of Applied Physics, Melbourne Technical College.

follows that the tracks of all particles recoiling at an angle θ, in a laboratory frame of reference, will be uniformly distributed around the surface of a cone of semi-angle θ, the axis, PO, being the incident direction. The slant-height of the cone is the range, L, in the emulsion for a recoil particle of this particular energy.



(a) (b)

Fig. 1

Suppose that the upper surface of the emulsion cuts the base of the cone along AB and that ψ is the angle AOB. From Figure 1 (*b*), which is portion of Figure 1 (*a*), we find

$$AO \cos \varphi = OP \sin \varphi + PK = L \cos \theta \sin \varphi + z,$$

that is

$$AO = \frac{L \cos \theta \sin \varphi + z}{\cos \varphi}.$$

Hence

$$\cos \psi = \frac{OA}{OB} = \frac{L \cos \theta \sin \varphi + z}{L \cos \varphi \sin \theta}. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

The ratio ψ/π gives the fractional loss through the upper surface for recoils of angle θ starting at a depth z below the surface. As is evident from Figure 2,

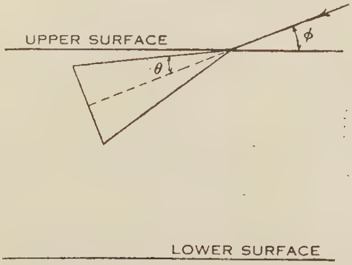


Fig. 2

for values of θ less than φ there is no upper surface loss since the cone does not intersect the surface. Considering the case $\theta > \varphi$ we see that ψ has its maximum value ψ_1 when P is in the surface and hence by putting $z=0$ in (**1**) we find

$$\psi_1 = \arccos(\tan \varphi / \tan \theta). \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

As the depth of origin of a recoil increases the loss fraction decreases from ψ_1/π at the surface to zero at a depth z given by

$$z_1 = L \sin(\theta - \varphi),$$

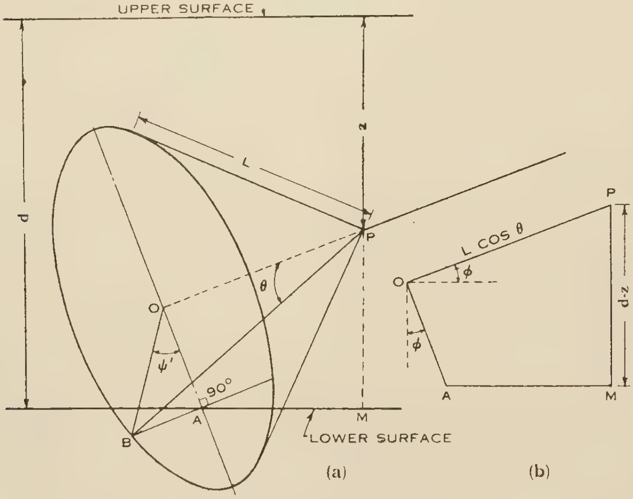
obtained by inserting $z = z_1$, and $\psi = 0$ in equation (**1**).

If N is the actual number of recoils of angle θ per unit thickness of the emulsion, then the number arising in a layer of thickness dz, at a depth z, and which escape through the surface

$$= \frac{\psi}{\pi} N\,dz.$$

Hence, on using the substitutions

$$a = (L \cos\varphi \sin\theta)^{-1} \text{ and } b = \tan\varphi / \tan\theta,$$



(a) (b)

Fig. 3

we have

$$\text{Total top surface loss} = \frac{N}{\pi}\int_0^{z_1} \psi\,dz,$$

$$= \frac{N}{\pi}\int_0^{z_1} \arccos(az+b)\,dz,$$

$$= \frac{N}{a\pi}\Big[\psi\cos\psi - \sin\psi\Big]_{\psi_1}^{0},$$

$$= \frac{N}{a\pi}(\sin\psi_1 - \psi_1\cos\psi_1). \quad \ldots\ldots\ldots \quad (\mathbf{3})$$

(*b*) *Loss Through the Lower Surface*

When the depth of the collision is such that the recoil cone cuts the lower surface of the emulsion as in Figure 3 (*a*) (part of which is redrawn in Fig. 3 (*b*)), then

$$OA = \frac{d - z - L\cos\theta\sin\varphi}{\cos\varphi},$$

where d is the thickness of the emulsion. Hence, denoting the angle AOB by ψ', we have

$$\cos\psi'=\frac{OA}{OB}=\frac{d-z-L\cos\theta\sin\varphi}{L\sin\theta\cos\varphi}. \quad \ldots\ldots\ldots\ldots (4)$$

Since, as is well known (Powell *et al.* 1946; Martin 1949), the emulsion shrinks considerably in thickness during the development and fixing procedure, it is important to note that d is the emulsion thickness before development. The magnitude of the shrinkage must be known since the calculation of θ for a track involves the depth measurement of a point on the track.

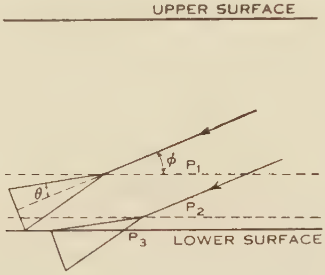


Fig. 4

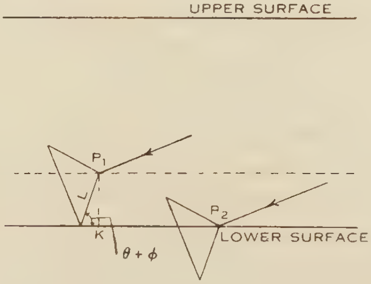


Fig. 5

It is clear from Figure 4 that for θ less than φ, some tracks in the slice P_1P_2 are not counted and there are no acceptable tracks at all in the layer P_2P_3 below this region. On making the substitution $b_1=ad-b$, the loss from P_1P_2 is given by

$$\begin{aligned}\frac{N}{\pi}\int_{\psi'=\pi}^{\psi'=0}\psi'dz&=\frac{N}{\pi}\int_{\psi'=\pi}^{\psi'=0}\arccos\,(b_1-az)dz,\\&=\frac{N}{a\pi}\Big[-\psi'\cos\psi'+\sin\psi'\Big]_{\pi}^{0},\\&=\frac{N}{a},\\&=NL\cos\varphi\sin\theta. \quad \ldots\ldots\ldots\ldots (5)\end{aligned}$$

Since the thickness of P_2P_3 is $L\sin(\varphi-\theta)$ the contribution of this portion to the lower surface loss is

$$NL\sin(\varphi-\theta). \quad \ldots\ldots\ldots\ldots (6)$$

Thus by adding (5) and (6) we find the total lower surface loss for scattering angles $\theta<\varphi$ to be given by

$$NL\cos\theta\sin\varphi. \quad \ldots\ldots\ldots\ldots (7)$$

When $\theta>\varphi$ (Fig. 5), ψ' varies from 0 for a collision at a height P_1K $(=L\sin(\varphi+\theta))$ above the bottom surface to a value

$$\psi'_1=\arccos(-\tan\varphi/\tan\theta) \quad \ldots\ldots\ldots\ldots (8)$$

when the recoil originates in the lower surface.

From (2) and (8) it follows that

$$\psi'_1=\pi-\psi_1,$$

and hence for θ greater than φ the escape-loss through the lower surface is

$$\frac{N}{\pi}\int_{\psi'=0}^{\psi'=\pi-\psi_1}\arccos(b_1-az)dz=\frac{N}{a\pi}\{(\pi-\psi_1)\cos\psi_1+\sin\psi_1\}. \quad \ldots\ldots (9)$$

(*c*) *Correction Factors*

Let N_θ be the number of tracks of angle θ which have been accepted for measurement. For all values of θ we have

$$\begin{array}{c}\text{Total number}\\\text{of tracks}\end{array}=\begin{array}{c}\text{Number of}\\\text{measured tracks}\end{array}+\begin{array}{c}\text{upper surface}\\\text{loss}\end{array}+\begin{array}{c}\text{lower surface}\\\text{loss}\end{array} \quad \ldots\ldots\ldots\ldots (10)$$

For θ less than φ this equation leads to

$$Nd=N_\theta+0+NL\cos\theta\sin\varphi,$$

and therefore

$$N=\frac{N_\theta}{d-L\sin\varphi\cos\theta}.$$

Hence the total number, Nd, of recoils of angle θ in the emulsion is given by

$$\frac{N_\theta d}{d-L\sin\varphi\cos\theta}$$

and thus the correction factor c_1 for tracks of this class is given by

$$c_1=\frac{d}{d-L\sin\varphi\cos\theta}. \quad \ldots\ldots\ldots\ldots (11)$$

When θ is larger than φ, on using (3) and (9) to form the equation expressed by (10) there results

$$Nd=N_\theta+\frac{N}{a\pi}\{\sin\psi_1-\psi_1\cos\psi_1\}+\frac{N}{a\pi}\{(\pi-\psi_1)\cos\psi_1+\sin\psi_1\},$$

from which it follows that the correction factor c_2 for θ greater than φ is given by

$$c_2=\frac{\pi d}{\pi d-L\cos\varphi\sin\theta\,[2\sin\psi_1-2\psi_1\cos\psi_1+\pi\cos\psi_1]}. \quad \ldots (12)$$

Since both equations (11) and (12) can be expressed in the form

$$\frac{\pi d/L}{\pi d/L-f(\theta,\varphi)},$$

where for $\theta<\varphi$

$$f(\theta,\varphi)=\pi\sin\varphi\cos\theta,$$

and for $\theta>\varphi$

$$f(\theta,\varphi)=\cos\varphi\sin\theta\,[2\sin\psi_1-2\psi_1\cos\psi_1+\pi\cos\psi_1],$$

two sets of curves can be drawn from which the correction factor can be obtained for any angle of incidence and for any given emulsion thickness. In Figure 6 (*a*), *f*(θ,φ) is plotted against θ for various values of φ whilst in Figure 6 (*b*) the loss factor is plotted against the recoil track length expressed as a fraction of the thickness of the emulsion. The appropriate value of *f*(θ,φ), read from Figure 6 (*a*), is used to find *c* from Figure 6 (*b*) for any given track length.

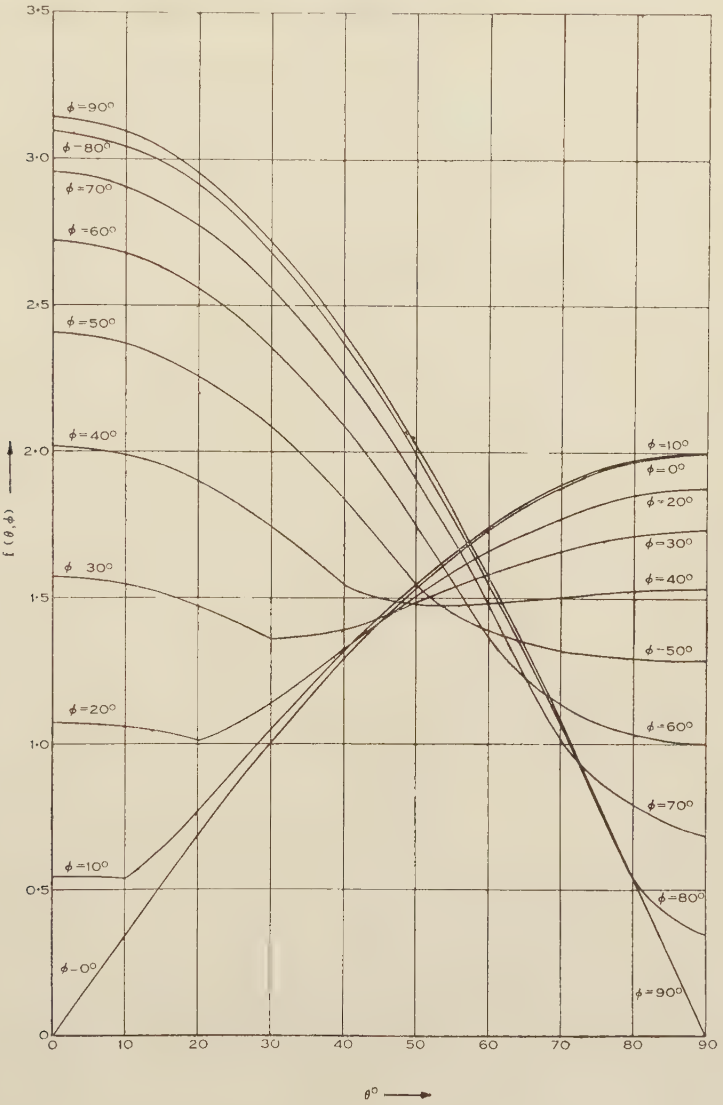


Fig. 6 (*a*)

Figure 6 (b) shows that c increases rapidly with increasing values of both L/d and $f(\theta,\varphi)$ and hence if c is to be known with reasonable accuracy each value of L/d used in a particular experiment should be on the flatter part of the appropriate curve. The thicker the emulsion used the more easily is this condition satisfied and the nearer to unity is the loss factor for any given case. For instance, in an n–p scattering experiment with D–D neutrons in which $\varphi=30°$, using data obtained from an energy-range curve (Powell and Occhialini 1947), we find that the factor for protons scattered in the head-on direction is 4·0 when the emulsion thickness is 50μ but is reduced to 1·3 when the emulsion thickness is increased to 200μ.

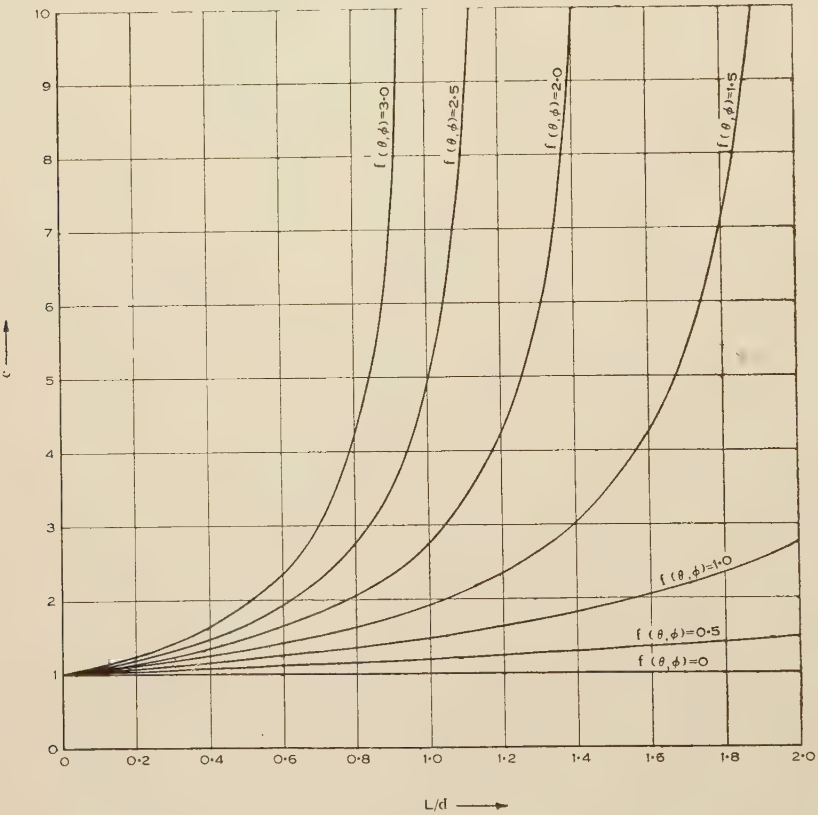


Fig. 6 (b)

On inserting $\varphi=0$ in (**12**) we obtain

$$c=\frac{d}{d-2L\sin\theta}, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{13})$$

and this gives the factor when the incident beam is parallel to the surface, this being the case for which Champion and Powell (1945) have given a graphical solution. It is important to note, however, that since these authors did not accept for measurement any track making an angle greater than 10° with the surface, equation (**13**) does not give the appropriate factor in their experiment for all angles of scattering. This dip criterion would have no effect on particles scattered at angles less than a certain angle β given by

$$\beta=\operatorname{artan}\left(\frac{\tan 10^\circ}{k}\right),$$

where the shrinkage ratio k, is defined by

$$k=\frac{\text{thickness after processing}}{\text{thickness before processing}}.$$

For those particles scattered at angles exceeding β it can be shown that the appropriate surface loss factor is given by

$$c=\frac{\pi d}{\pi d-2\operatorname{arsin}\left(\dfrac{\sqrt{\cos^2\beta-\cos^2\theta}}{\sin\theta}\right)+2L\sqrt{\cos^2\beta-\cos^2\theta}-2L\sin\theta}$$

III. LIMITATION OF THE FORMULAE

In making the calculations small-angle scattering has been neglected since it has been assumed that all the recoil tracks are rectilinear. Powell and Occhialini (1946) have given an empirical method for estimating the surface loss when small-angle scattering is taken into account and they concluded from measurements on proton tracks in half-tone emulsions of 150μ thickness that the effect of Coulomb scattering was to increase the surface loss. Their results showed, however, that the effect was not appreciable for protons of energy 9 MeV. or less, but that serious errors were introduced for protons of energy greater than this value. As these authors have pointed out it is to be expected that the increased Coulomb scattering in the nuclear research emulsions over that in half-tone emulsions would lower this energy limit when nuclear emulsions were used but no data are available on this point. It must be remembered, however, that since in Powell and Occhialini's experiments the incident beam of particles was parallel to the emulsion surface all, except head-on recoils, would be directed towards either the upper or lower surface and on this account it might be expected that the effect due to small-angle scattering would be more important for this case than for that when the incident beam is inclined to the surface. Taking these considerations into account and also on the basis of experience with proton, deuteron, and α-particle tracks, the author believes that the formulae developed in the present paper give reliable results to particle energies at least as high as 5 MeV.

IV. ACKNOWLEDGMENTS

The author wishes to acknowledge his indebtedness to Dr. E. H. S. Burhop, University College, London, for his helpful correspondence ; to his colleagues Mr. J. M. Allen and Mr. A. K. Connor, Melbourne Technical College, for advice ; and to Mr. A. Mott and Mr. C. Chamberlain for help in preparing the diagrams.

V. REFERENCES

BARSCHALL, H. H., and KANNER, M. H. (1940).—*Phys. Rev.* **58** : 590.

CHAMPION, F. C., and POWELL, C. F. (1945).—*Proc. Roy. Soc.* A **183** : 64.

DEE, P. L., and GILBERT, C. W. (1937).—*Proc. Roy. Soc.* A **163** : 265.

MARTIN, S. L. (1949).—*Aust. J. Sci. Res.* A **2** : 389.

MARTIN, S. L., BURHOP, E. H. S., ALCOCK, C. B., and BOYD, R. L. F. (1950).—*Proc. Phys. Soc.* A **63** : 884.

POWELL, C. F., and OCCHIALINI, G. P. S. (1946).—" Report of Cambridge Conference on Fundamental Particles and Low Temperatures." Vol. 1, p. 150. (The Physical Society : London.)

POWELL, C. F., and OCCHIALINI, G. P. S. (1947).—" Nuclear Physics in Photographs." Fig. 4, p. 34. (Clarendon Press : Oxford.)

POWELL, C. F., OCCHIALINI, G. P. S., LIVESEY, D. L., and CHILTON, L. V. (1946).—*J. Sci. Instrum.* **23** : 102.

RICHARDS, H. T. (1941).—*Phys. Rev.* **59** : 796.

THE TEMPERATURE DEPENDENCE OF THE THERMAL CONDUCTIVITY OF AIR

By W. G. KANNULUIK* and E. H. CARMAN*

[*Manuscript received April 5, 1951*]

Summary

The thermal conductivity of air is determined in absolute measure between —183 and 218 °C. by the " thick-wire " variant of the " hot-wire " method. The apparatus consists of a platinum wire $11\cdot 6$ cm. long and $1\cdot 5$ mm. in diameter mounted in a 90 per cent. Pt 10 per cent. Ir tube of 7 mm. internal diameter. The temperature variation of the thermal conductivity of air can be represented with considerable accuracy by the quadratic formula

$$k_t=5\cdot 75\times 10^{-5}(1+0\cdot 00317t-0\cdot 0000021t^2).$$

The present paper suggests a value for the thermal conductivity of air at 0 °C. from a consideration of the data which have been obtained by various hot-wire methods.

I. INTRODUCTION

The experimental investigation of the variation of the thermal conductivity of gases with temperature has been the subject of but few full-scale researches. Eucken (1911, 1913) studied the temperature variation of the conductivity of a number of gases between —190 and 100 °C. He used a " hot-wire " method with a compensating tube, and obtained results relative to air at 0 °C.

Recently, Taylor and Johnston (1946) and Johnston and Grilly (1946) have obtained absolute values of the conductivity of several gases at a large number of temperatures between —190 and 100 °C. In those two investigations the form of hot-wire method introduced by Schleiermacher (1888) was used.

The authors of the present paper felt that a knowledge of the temperature variation of the conductivity is of sufficient interest to warrant a further careful investigation. In this laboratory a " thick-wire " variant of the hot-wire method has been used successfully over a period of years for conductivity measurements at 0 °C. It was decided to adopt this method for measurements over the temperature range —183 to 218 °C. Some exploratory measurements on carbon dioxide (Kannuluik and Law 1947) between $-78\cdot 5$ and 100 °C. by the thick-wire apparatus had already proved its suitability for this work.

II. DESCRIPTION OF THE APPARATUS

The conductivity apparatus used in the present work was designed for accurate absolute measurements and for use over a range of temperatures. For the hot-wire a platinum wire $1\cdot 5$ mm. in diameter was used. This was mounted in a thin-walled 90 per cent. Pt 10 per cent. Ir tube $11\cdot 6$ mm. long and

---

* Physics Department, University of Melbourne.

7 mm. in internal diameter. The alloy was preferred to pure platinum for the tube in order to obtain greater mechanical strength. The choice of materials also assures that there is practically no differential expansion of the wire and the tube. In design the present apparatus is similar to that used in earlier work with the thick-wire method (Kannuluik and Law 1947 ; Kannuluik and Donald 1950).

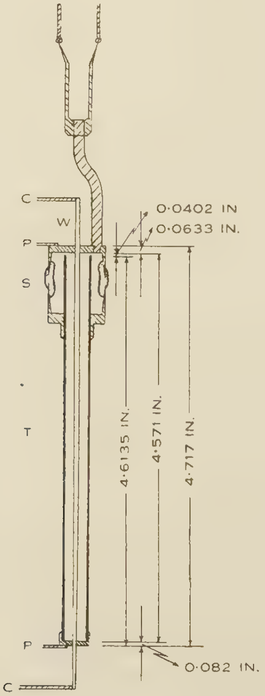


Fig. 1.—The conductivity apparatus.

Particulars of the apparatus are shown in Figure 1, in which *W* is the platinum wire, *T* the platinum-iridium tube, and *S* a double copper glass seal which serves to insulate the wire electrically from the tube. The leads, labelled *C* and *P*, are current and potential leads respectively. In assembling the parts of the apparatus silver soldering was used throughout.

(*a*) *Absence of Convection and the Temperature " Jump " Effect*

The above apparatus is entirely free from the effects of convection. Also the temperature " jump " at the wire surface is without appreciable effect until the gas pressure is reduced to a few cm. of mercury. This is true for air except at the upper limit of the temperature range (218 °C.) where there does appear to be a very small effect even at ordinary pressures. The conditions to be complied with for absence of convection have been analysed in a previous paper (Kannuluik and Donald 1950).

(*b*) *Dimensions of the Apparatus*

The special facilities of the Metrology Section of the Defence Research Laboratory, Maribyrnong, were made available through the kindness of Mr. K. C. Lang who determined for us the internal diameter of the tube at short intervals along it and also the diameter of the platinum wire.

The following are the principal dimensions of the apparatus :

Diameter of the wire	0·1465 ± 0·0001 cm.
Length of the wire	11·610 ± 0·005 cm.
Mean internal diameter of the tube ..	0·7169 ± 0·0005 cm.

III. Theory

A brief outline of the theory used in the present work will be given. A detailed account has already been given in an earlier paper (Kannuluik and Law 1947). The differential equation for a steady flow of heat along a wire mounted axially in a tube filled with gas and maintained everywhere at zero temperature is

$$\lambda\pi b^2\frac{\partial^2\theta}{\partial z^2}+2\pi bk\left(\frac{\partial\theta}{\partial r}\right)_{r=b}+\frac{I^2\rho_0}{J}(1+\alpha\theta)=0, \quad \dots\dots\dots\dots\dots\dots \quad (1)$$

where λ is the thermal conductivity of the wire, k the thermal conductivity of the gas, $\rho=\rho_0(1+\alpha\theta)$ the resistance per unit length of the wire at a temperature θ, and b is the radius of the wire.

For the purposes of calculation, it is more convenient to assume a strictly radial flow of heat through the gas, the second term of the above equation is then constant and the equation has the solution

$$\left.\begin{aligned}\left(\frac{1}{\beta l}\right)^2\left(1-\frac{\tanh\beta l}{\beta l}\right)&=\frac{2\pi b^2\lambda J(R-R_0)}{R_0{}^2I^2\alpha l},\\ \text{where}\qquad \beta^2&=\frac{2h}{b\lambda}-\frac{I^2R_0\alpha}{2\pi b^2J\lambda l},\end{aligned}\right\}\quad \dots\dots\dots \quad (2)$$

and

$$h=\frac{k}{b\ln\dfrac{a}{b}}, \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \quad (3)$$

R (ohm) is the resistance of the wire at its mean temperature when a current I (ampere) is flowing,

R_0 is the true resistance of the wire at $\theta=0$,

α is the temperature coefficient of the resistance,

J is the mechanical equivalent of heat,

$2l$ is the length and $2b$ the diameter of the wire,

$2a$ is the internal diameter of the tube,

h is the external conductivity.

When the hot-wire apparatus is highly evacuated, the only radial heat loss is that due to radiation. The external conductivity h may then be replaced by

h_R, where $h_R << h$ for moderate temperatures. It can be shown that equation (2) then reduces to

$$\lambda=\frac{1}{6}\,\frac{RR_0I^2\alpha l}{\pi b^2J(R-R_0)}\left(1+\frac{1}{30}\,\frac{R_0I^2\alpha l}{\pi b^2J\lambda}\right)\left(1-\frac{4}{5}\,\frac{h_Rl^2}{b\lambda}\right). \qquad (4)$$

The value of λ obtained from (4) may be substituted in equation (2). This may be solved for h and finally the thermal conductivity k obtained from (3).

Two corrections are applied to the value of h obtained from equation (2): (i) for radiation, and (ii) for departure from strict radial-flow conditions. The first of these is of course negligible at the lower end of the temperature range but becomes appreciable at the higher temperatures used in this work. The second correction is small and amounts to a reduction in h of the order of 1 per cent.

IV. CALIBRATION OF THE APPARATUS

The authors preferred to reproduce a number of standard and substandard temperatures on the International Temperature Scale rather than to use a thermostat. The principal reasons for this choice were, firstly, the simplicity of the ancillary apparatus, and secondly, an accurate knowledge of the temperatures at which the conductivity measurements were made, was assured.

The following standard and substandard temperatures were employed:

(i) The temperature of equilibrium between liquid oxygen and its vapour (−182·97 °C.).
(ii) The temperature of equilibrium between solid carbon dioxide and its vapour (−78·5 °C.).
(iii) The melting point of ice (0 °C.).
(iv) The boiling point of water (100 °C.).
(v) The temperature of equilibrium between liquid naphthalene and its vapour (218·0 °C.).

Remarks on the reproduction of the above temperatures

−182·97 °C.—The conductivity apparatus was immersed in a vacuum flask, 18 in. long and $2\frac{1}{2}$ in. in diameter, filled with liquid oxygen. The temperature at the mid-point of the apparatus was measured at frequent intervals with a vapour-pressure thermometer in the bulb of which was condensed pure oxygen generated by heating potassium permanganate.

−78·5 °C.—A method of maintaining a "dry-ice" bath to give this temperature was developed in this laboratory a few years ago and an account of the method has been published (Kannuluik and Law 1946). This technique was used in the present work.

100 °C.—A hypsometer of simple design electrically heated by means of direct current gave good results at the steam point.

218 °C.—A simple type of hypsometer heated by means of a bunsen flame was used with satisfactory results at the naphthalene point.

The calibration of the apparatus involves the determination of the resistance, temperature coefficient of resistance, and thermal conductivity of the wire at each of the five temperatures. All these data are obtained from a set of observations of the resistance $\bar{R}$ corresponding to a series of heating currents I at each of the temperatures with the tube highly evacuated. In the present work an oil-diffusion pump was used to produce a vacuum better than 10^{-5} mm. of mercury.

The resistance R_t at each of the temperatures is obtained by extrapolating the accurate linear relation between $1/R$ and I^2 to zero current.

In accordance with the standard procedure in platinum thermometry the values of R_t for the range 0 to 218 °C. were fitted to a quadratic formula of the form

$$R_t = R_0(1 + At + Bt^2).$$

For the range below 0 °C. the formula

$$R_t = R_0\{1 + At + Bt^2 + C(t-100)t^3\}$$

was used ; the values of the three constants A, B, and C obtained by this means were :

$$A = 0\cdot0039570\ ;\quad B = -5\cdot872 \times 10^{-7}\ ;\quad C = -3\cdot853 \times 10^{-12}.$$

The values of R_t determined at the five different temperatures are given in the first column of Table 1.

The temperature coefficients of resistance,

$$\left[\alpha_t = \frac{1}{R_t}\left(\frac{dR_t}{dt}\right)_t\right],$$

are obtained by differentiating the above formulae for R_t. The values of α_t are given in the second column of Table 1.

TABLE 1

CALIBRATION DATA

t (°C.)	R_t	α_t	λ (cal. cm.$^{-1}$ sec.$^{-1}$ °C.$^{-1}$)
—182·97	0·00169707	0·01721	0·1657±0·0005
—78·5	0·00466105	0·005927	0·1623±0·0004
0	0·00679787	0·003957	0·1679±0·0003
100	0·00944787	0·002763	0·1721±0·0003
218·0	0·0124722	0·002017	0·1735±0·0002

By using the data of the second and third columns of Table 1, the thermal conductivity λ of the wire at the various temperatures is calculated. The values of λ given in the fourth column of Table 1 were calculated by means of formula (4) ; each value is the mean of six determinations of λ for a range of values of the current I.

V. EXPERIMENTAL RESULTS

Air for use in the measurements was dried by passing it over calcium chloride and phosphorus pentoxide. The small amount of carbon dioxide present was not removed.

TABLE 2

THE THERMAL CONDUCTIVITY OF AIR

—182·97 °C.		—78·5 °C.		0 °C.		100 °C.		218·0 °C.	
p (cm.)	k* × 10^5	p (cm.)	$k \times 10^5$	p (cm.)	$k \times 10^5$	p (cm.)	$k \times 10^5$	p (cm.)	$k \times 10^5$
76·20	2·013	76·05	4·315	65·26	5·737	75·40	7·453	77·43	9·151
75·69	2·008	51·04	4·309	52·96	5·738	57·10	7·447	63·62	9·136
57·82	1·991	40·86	4·253	43·07	5·742	46·88	7·441	42·35	9·117
47·11	2·004	30·78	4·263	33·08	5·743	36·58	7·431	39·77	9·105
36·87	1·971	20·17	4·262	23·15	5·739	27·06	7·450	27·31	9·067
26·45	2·023	10·15	4·278	13·37	5·746	16·60	7·459	15·25	8·995
16·01	2·024	6·81	4·282	6·93	5·747	6·30	7·467	9·36	9·020
6·24	2·009	3·97	4·237	2·88	5·727	2·35	7·431	3·39	8·884
3·26	2·000	1·41	4·233	1·24	5·667	1·26	7·385	1·16	8·598
1·66	2·000					0·48	7·205		
0·48	2·010								
2·01±0·01 ; (0·6°)†		4·27±0·02 ; (1·2°)		5·741±0·003 ; (1·4°)		7·45±0·01 ; (1·7°)		9·15 ; (1·8°)	

* Unit of k : cal. cm.$^{-1}$ sec.$^{-1}$ °C.$^{-1}$.

† At the foot of the columns are mean values of k from appropriate pressure ranges (see text) ; the numbers in brackets indicate the mean rise in temperature of the air in the tube.

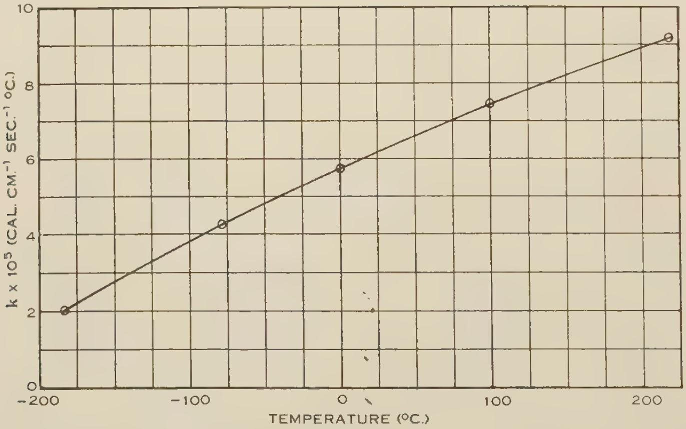


Fig. 2.—Variation of the thermal conductivity of air between —183 and 218 °C.

The values of the thermal conductivity and the corresponding pressure of the gas are given for each of the five temperatures in Table 2. At the foot of the columns are given the mean values of k calculated for the range of pressures over which k is constant except in the case of the highest temperature (218·0 °C.). The approximate pressure range is 6 to 70 cm. of mercury. The value of k for 218·0 °C. was corrected for the effect of a small temperature jump by plotting

p/k against the pressure p and deducing the conductivity from the slope of the resulting straight line. At a pressure of 40 cm. this correction amounts to 0·5 per cent.

The mean values of k given in Table 2 are shown plotted against the temperature in Figure 2.

VI. Discussion of the Experimental Results

(*a*) *The Independence of Conductivity on the Pressure*

Owing to the absence of convection and the temperature jump effect in the thick-wire variant of the hot-wire method, it is possible to obtain an exact verification of Maxwell's prediction from the kinetic theory of the independence of the conductivity on the pressure. For example, in Table 2 the conductivity of air at 0 °C. is constant over the pressure range 6 to 76 cm. of mercury to better than one part in 1500. This probably represents the most accurate verification of Maxwell's prediction that has been obtained.

(*b*) *The Thermal Conductivity of Air at 0 °C.*

The thermal conductivity k_0 of air at 0 °C. is a specific constant of considerable importance. The question of the exact value to be assigned to it is still undecided. Suggested values of $k_0 \times 10^5$ obtained by some form of the hot-wire apparatus range from 5·85 to 5·66. The values of k_0 which have been obtained in this laboratory by the thick-wire variant using a variety of wires mounted in

Table 3

VALUES OF k_0 DETERMINED WITH THE THICK-WIRE METHOD

$k_0 \times 10^5$	Form of Apparatus	Reference
5·75	Copper wire 0·6 mm. diameter surrounded by a copper tube 16 cm. long and 1 cm. diameter	Kannuluik and Martin (1934)
5·76	Copper wire (as above) surrounded by a copper tube 10 cm. long and 1·6 cm. diameter	Kannuluik and Martin (1934)
5·76	Platinum wire 1·5 mm. diameter in a stainless steel tube 1·3 cm. diameter	Kannuluik and Martin (1934)
5·74	Two coaxial nickel tubes	Kannuluik and Donald (1950)
5·74	Platinum wire in platinum-iridium tube	Present work

tubes of different dimensions form a concordant set. The relevant data are given in Table 3. The mean value of k_0 is $5\cdot75 \pm 0\cdot1 \times 10^{-5}$ cal. cm.$^{-1}$ sec.$^{-1}$ °C.$^{-1}$.

Data for the same constant obtained by other observers using hot-wire apparatus are given in Table 4. Although the spread of the data is considerably greater the mean of the values listed is $k_0 = 5\cdot74 \pm 0\cdot1 \times 10^{-5}$ which is practically the same as the mean of the values obtained in this Laboratory. It appears much more reasonable to accept a mean value of $k_0 = 5\cdot74 \times 10^{-5}$ as better representing the value of the conductivity of air at 0 °C. by the hot-wire method than either of the extreme values $5\cdot66 \times 10^{-5}$ or $5\cdot85 \times 10^{-5}$. Furthermore a determination of k_0 has been made by Hercus and Sutherland (1934) using an entirely different method not subject to any of the inherent errors to which the hot-wire method is liable. Hercus and Sutherland used the parallel-plate method and obtained for air at 0 °C. the value $k_0 = 5\cdot72 \times 10^{-5}$. This value is sufficiently close to the mean of the hot-wire values ($5\cdot74 \times 10^{-5}$) to confirm the correctness of the value deduced from the hot-wire experiments.

TABLE 4

VALUES OF k_0 DETERMINED WITH OTHER THICK-WIRE METHODS

$k_0 \times 10^5$	Reference
5·68	Weber (1917)
5·74	Weber (1927)
5·85	Gregory and Archer (1933)
5·81	Milverton (1934)
5·66	Taylor and Johnston (1946)

(c) *The Dependence of the Conductivity of Air on the Temperature*

The values of the conductivity k appearing in Table 2 have been corrected for radiation, for departure from strict radial flow, and reduced to the appropriate standard and substandard temperatures by assuming a temperature coefficient of $k = 0\cdot003$. For the values at $-182\cdot97$ °C. the radiation correction is negligible. At $-78\cdot5$ °C. the radiation amounts to an equivalent conductivity of $0\cdot012 \times 10^{-5}$ and increases to $0\cdot046 \times 10^{-5}$ at 0 °C., $0\cdot162 \times 10^{-5}$ at 100 °C., and $0\cdot486 \times 10^{-5}$ at 218·0 °C. A 1 per cent. reduction in each value was applied for the departure from radial flow.

Table 5 contains values of the constant ε calculated from the well-known formula of the kinetic theory, viz.:

$$k = \varepsilon \eta c_v.$$

To obtain the values of ε at the different temperatures it is necessary to know the corresponding values of η. These values of η have been calculated from the Sutherland (1893) formula:

$$\eta = \frac{K\theta^{3/2}}{\theta + C},$$

TABLE 5

VALUES OF $k/(\eta c_v)$

t (° C.)	$k\times10^5$ (cal. cm.$^{-1}$sec.$^{-1}$ °C.$^{-1}$)	$\eta\times10^7$ (c.g.s.)	ε*
—182·97	2·01	618·6	1·92
—78·5	4·27	1306	1·92
0	5·74	1735	1·95
100	7·45	2206	1·99
218·0	9·15	2684	2·02

* It is well known that Eucken has deduced the relation $\varepsilon=\frac{1}{4}(9\gamma-5)$ from theoretical considerations, and that his value of ε for air is 1·90 taking $\gamma=1\cdot40$. In the above table c_v was taken to be 0·1693.

where θ represents the absolute temperature, and K and C are numerical constants. The values of K and C (given in Kaye and Laby's " Physical and Chemical Constants ", p. 44) are :

$$K=1\cdot5\times10^{-7},\text{ and } C=117.$$

TABLE 6

VALUES OF k/k_0

t (°C.)	Eucken (1911)	Taylor and Johnston (1946)	Present Work
--182·97	0·354	0·350	0·350
—78·5	0·752	0·751	0·744
0	1·00	1·00	1·000
100	1·27	1·29	1·30
218·0			1·59

The values for k obtained in this work cannot be fitted to a Sutherland formula. However, over the whole range of temperatures the observations fit the following quadratic formula :

$$k_t=5\cdot75\times10^{-5}(1+0\cdot00317t-0\cdot0000021t^2).*$$

A comparison of the relative conductivity k_t/k_0 in the temperature range —183 to 218 °C. for the earlier work of Eucken (1911) and for that of Taylor and Johnston (1946) with the present work is given in Table 6.

(d) *The Temperature Coefficient of the Thermal Conductivity of Air at 0 °C.*

From the quadratic formula quoted above, the temperature coefficient of k is 0·00317. In Table 7 are given some other determinations of this quantity by observers whose observations of k were obtained over a wide range of temperature.

---

* This formula was obtained by the method of least squares.

Table 7

TEMPERATURE COEFFICIENTS OF AIR AT 0 °C.

Temperature Coefficient of k at 0 °C.	Temperature Range of Observations (0 °C.)	Reference
0·00298	—191–212	International Critical Tables*
0·00271	—190–100	Eucken (1911)
0·00291	—190–100	Taylor and Johnston (1946)
0·00317	—183–218	Present work†

* Computed from data of various investigators over the period 1878–1913.

† This value is greater than the mean temperature coefficient between 0 and 100 °C. which is 0·00298.

VII. Acknowledgments

The authors wish to record their indebtedness to Professor L. H. Martin for the facilities provided, and to thank him for his continued interest in the work.

VIII. References

Eucken, A. (1911).—*Phys. Z.* **12** : 1101.

Eucken, A. (1913).—*Phys. Z.* **14** : 324.

Gregory, H. S., and Archer, C. T. (1933).—*Phil. Mag.* **15** : 301.

Hercus, E. O., and Sutherland, D. M. (1934).—*Proc. Roy. Soc.* A **145** : 599.

Johnston, H. L., and Grilly, E. R. (1946).—*J. Chem. Phys.* **14** : 233.

Kannuluik, W. G., and Donald, Heather B. (1950).—*Aust. J. Sci. Res.* A **3** . 417.

Kannuluik, W. G., and Law, P. G. (1946).—*J. Sci. Instrum.* **23** : 154.

Kannuluik, W. G., and Law, P. G. (1947).—*Proc. Roy. Soc. Vict.* **58** : 142.

Kannuluik, W. G., and Martin, L. H. (1934).—*Proc. Roy. Soc.* A **144** : 496.

Milverton, S. W. (1934).—*Phil. Mag.* **17** : 414.

Schleiermacher, A. (1888).—*Ann. Phys. Lpz.* **34** : 623.

Sutherland, W. (1893).—*Phil. Mag.* (5) **36** : 507.

Taylor, W. J., and Johnston, H. L. (1946).—*J. Chem. Phys.* **14** : 219.

Weber, S. (1917).—*Ann. Phys. Lpz.* **54** : 325.

Weber, S. (1927).—*Ann. Phys. Lpz.* **82** : 479.

A SURVEY OF THE STRESS BETWEEN THE OCEAN AND ATMOSPHERE

By C. H. B. Priestley*

[*Manuscript received February* 9, 1951]

Summary

A survey is made of the mean east-west component of stress between the ocean and atmosphere for each season and each ocean, and for the mean of all oceans. The averaging is carried out for each 5° latitude zone from 55 °N. to 55 °S.

The results provide a first approximation to the total torque about the Earth's axis, on the atmosphere as a whole, in the various latitude belts. From them is derived a table showing the distribution of total northward flux of atmospheric angular momentum for the annual mean and the four seasons individually, allowing for the changes occurring during spring and autumn.

A brief comparative discussion of both the stress and flux results is given.

I. Introduction

Mechanical stresses between the atmosphere and the ocean arise as a result of the relative movement in the boundary layers of the two media close to the sea surface ; primarily, since the atmosphere is the more mobile of the two, they may be regarded as an effect of wind. In practice it is convenient to make some distinction between wind and stress patterns which are relatively short-lived and localized and those which are, in the case of the atmosphere, world-wide and, in the case of the ocean, at least ocean-wide in their operation.

In the first category are the effects which occur on the scale of the major transitory wind systems such as depressions and fronts, tropical storms, and moving anticyclones. The stress is then significant to the atmosphere in that it represents a major destructive influence to the wind and pressure system, not only by the direct process of kinetic energy dissipation but indirectly by bringing about a surface airflow from high to low pressure and so tending to destroy the pressure contrast : and it is equally significant to the ocean in being an agency for the generation of waves and swell and the smaller wind-driven surface currents and induced motions, e.g. upwelling.

In this paper we shall rather be concerned with the more widespread and semi-permanent features of the circulation and stress pattern which are revealed by the mean monthly and annual charts. It is known that the prevailing surface wind has an easterly component in low and a westerly component in middle latitudes : in these zones therefore the torque of the mean stress acts respectively as a source and a sink of eastward atmospheric angular momentum about the Earth's axis. A study of these torques is necessary to an understanding of the changes in the zonal circulation of the atmosphere, its distribution with latitude,

---

* Section of Meteorological Physics, C.S.I.R.O., Melbourne.

and the closely allied changes in weather type. Nevertheless the stress distribution has not yet been subjected to a systematic quantitative survey. The main purpose of this paper is to make a first attempt to fill this gap by assessing the mean stresses season by season and latitude by latitude as far as available data warrant, and by inferring therefrom the total seasonal flow of angular momentum between the latitudes.

The data to which recourse must be made consist of wind measurements and observations, and it must be conceded that the formula representing the relation between wind and stress is an approximation. Furthermore the ocean covers most of the Earth's surface but not the whole, and the extension of the work to the land areas would encounter certain rather forbidding difficulties. The figures derived here, then, may lack some measure of accuracy and of representativeness for the atmosphere as a whole. But despite the present unavoidable limitations it is thought that this first attempt will satisfy an urgent need, besides serving as an incentive to more detailed and accurate studies. The parallel work on the energy balance has developed in just this way, the poleward flow of energy having been discussed by several workers over many years with progressively increasing detail and accuracy.

In the atmosphere we are basically concerned with the maintenance of the general circulation, and since there are no lateral boundaries the mean stress over all longitudes is of significance to the total circulation. The oceans, on the other hand, are interconnected only in certain latitudes, and in the majority of latitudes the average stress over all longitudes would have no physical significance in the converse problem of physical oceanography, i.e. the maintenance of the major wind-driven ocean surface currents. For this purpose, therefore, the data have been subdivided by oceans, with the separate evaluation over each ocean of the mean stress distributions.

II. The Wind-Stress Relation and its Application

In aerodynamic work it has been established that the resistance to flow past a solid body varies as the square of the velocity over a wide range of wind speeds for which the boundary layer can be regarded as fully rough. The case of flow over the oceans is more complicated since both media are mobile and the height and form of the waves, or roughness elements, will themselves depend strongly on the wind speed. Nevertheless it has been found (see, e.g. Sverdrup, Johnson, and Fleming 1946) that the stress τ and surface wind speed V can be roughly related in the form

$$\tau = \varkappa \rho V^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

over a fair range of wind speeds. ρ is air density and $\varkappa$ a number which is normally regarded as a constant, though determinations of its value by different workers at different times and places and in different conditions vary considerably. There is evidence, for example, that $\varkappa$ may increase quite abruptly at a wind speed of about 7 m./sec. (Munk 1947), but it cannot be said that a unique law of variation of $\varkappa$ with V has yet been established with any certainty.

Moreover, even if the determinations of τ had established some such accurate law, it would have to be borne in mind that the determinations have, of necessity, been made in relatively steady conditions, or confined waters, or in regions of relatively short (and therefore uniform) wind fetch. It seems to the writer that over the open oceans no unique relation between τ and V could be expected to apply exactly. The complicating factor is the variability of the wind. For a large region over which the wind is fairly constant and uniform, albeit strong, the surface current and wave form adjust themselves to the wind so as to present an effectively smaller degree of roughness than would obtain over a region where the wind, though of the same mean strength, is more variable in space and time. The bearing of this on the present problem is to imply that in a statistical sense an effectively higher value of $\varkappa$ would be applicable to the zone of surface westerlies than to the more constant and uniform easterlies of tropical latitudes : but present data do not indicate the magnitude of the effect. There is one saving grace in that the " wind observations " are in fact, for the main part, observations of the state of the sea so that the error in stress estimates from this cause may not be so great as would at first sight be supposed.

In undertaking a survey of the stress distribution a choice between two computing procedures is available. The first is to invoke some law for $\varkappa$ as a function of V. From the discussion above it is evident that any such law could only be tentative, and in view of the other uncertainties it is doubtful if such an elaboration would at present be warranted. The alternative adopted here is to rely on the approximating power of statistics and use, as a constant $\varkappa$ value, the mean of all determinations to date which according to Deacon (1950) is $0 \cdot 0013$. This may be assumed to give at least a good representation of the main features of the latitude-pattern and its seasonal variation. In so far as the main results of this first estimate are of a comparative nature, the method is quite adequate for the purpose in hand. Furthermore, from a brief discussion of the error inherent in this procedure, it is possible to assign its sense and order of magnitude so that, in applications such as that of Section VI where accuracy of absolute values is sought, it is possible to make some allowance and so achieve an accuracy comparable with that at present attainable by any other procedure.

Since $0 \cdot 0013$ is the mean of all determinations, irrespective of wind speed, the quadratic form of (1) implies that the mean stress will be underestimated if $\varkappa$ does, in fact, increase with increasing wind speed. The underestimation will be greatest in those regions where the stronger winds are relatively frequent ; that is, will be greater in the temperate than in the tropical oceans, the very high winds of tropical cyclones not being of sufficient frequency to reverse this contrast. The evidence discussed by Sverdrup, Johnson, and Fleming (1946) and Munk (1947), however, indicates that $\varkappa$ does not significantly exceed $0 \cdot 0026$ in the strongest wind so that the underestimation cannot be greater than 50 per cent. On this basis we conclude that all the estimates are between 50 and 100 per cent. of the true value, except possibly in regions where the mean estimated stress component is small, and not of great physical importance. These may be regarded as tolerable limits of accuracy particularly since it is only to the extent that the underestimation may vary, with latitude and season, within this range that any modifications might be required to the comparisons to be drawn here.

III. WORKING DETAILS

If f denotes the frequency of wind of strength v from direction θ, measured clockwise from north, the mean stress component $\bar{\tau}_x$ on the atmosphere in the direction from west to east is

$$\bar{\tau}_x = \varkappa\rho \Sigma f v^2 \sin \theta, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

neglecting as unimportant the variations of ρ with v and θ. The mean value of this quantity was determined from wind frequency roses, for a large number of years, over the oceans in 5° latitude-longitude areas, from 55° N. to 55° S., the results being averaged in 5° latitude zones for the Earth as a whole and sub-meaned for each ocean separately. The evaluations were made for each of the four months January, April, July, and October, and the means for the whole year also extracted. The whole task involved the working-up of some 4000 wind roses.

No account was taken of the fact that the mean value of ρ should vary with place and, to a smaller extent, with season, a value of $0 \cdot 00125$ g. cm.$^{-3}$ being used throughout the computation. The error involved is merged in the larger errors arising from other causes. It is felt that a more elaborate procedure in this respect will be warranted only when the wind-stress relation is more exactly known.

The source of information used for latitudes 55° N. to 30° S. was the British Admiralty publication entitled "Meteorological Charts of the World", the data being corrected up to 1943. From 30° to 55° S. material was taken from three charts published by the Hydrographic Office, U.S. Navy Department:

Pilot Charts of the South Atlantic Ocean, No. 2600 (1949),
Pilot Charts of the South Pacific Ocean, No. 2601 (1950),
Pilot Charts of the Indian Ocean, No. 2603 (1950).

No. 2601 was also used to make good a gap in the British Admiralty Charts in certain longitudes from 10° to 30° S.

The observations from 50 to 55° S. were further supplemented by information drawn from British Meteorological Office publications (M.O. 484, M.O.M. 402, 446, and 470). With the exception of these last-named, the wind roses showed only frequencies and mean speeds $\bar{v}$ from each octant, so that in effect equation (2) was approximated by

$$\bar{\tau}_x = \varkappa\rho \Sigma f_1 \bar{v}^2 \sin \theta, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

f_1 now denoting the total frequency from each octant. A considerable saving in computation labour was achieved, but it is clear that a consistent underestimate must thereby result.

The M.O. charts, however, gave frequencies from each octant in various wind speed ranges, allowing (2) to be more accurately evaluated. A comparison of the results obtained from (2) and (3) was therefore possible. From a number of wind roses chosen from all latitudes and longitudes at random, it was found that the value obtained from (2) exceeded that from (3) by an average of 25 per cent. between 30° N. and 30° S., and by 33 per cent. polewards of these latitudes. No significant variation in these factors was found as between hemispheres, between seasons, or within the zones delineated. They were therefore applied as constant factors to the whole of the material computed from (3). There remains some slight degree of underestimation due to the grouping into speed ranges of even the most detailed wind roses available, but this is not unduly large as compared with other errors and is common to all latitudes and seasons.

Tests between the various sources in overlapping regions revealed no inconsistency.

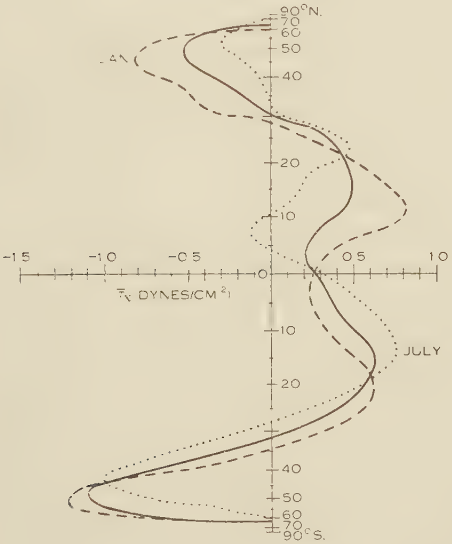


Fig. 1 (*a*)

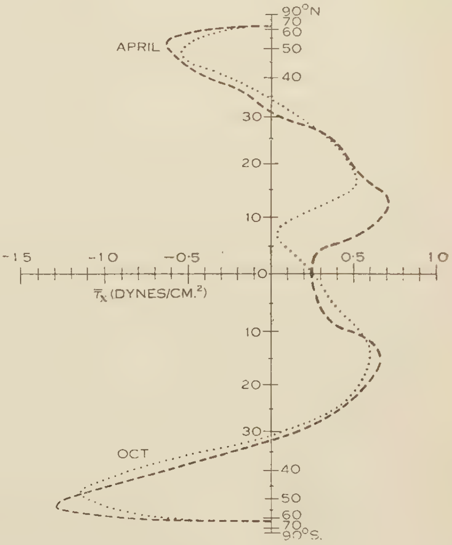


Fig. 1 (*b*)

Fig. 1 (*a*).—Mean stress distribution over all oceans for January (dashed), July (dotted), and annual (full line).

Fig. 1 (*b*).—Mean stress distribution over all oceans for April (dashed) and October (dotted).

The outcome of these computations was to obtain $\bar{\tau}_x$ (or more exactly $\overline{v^2 \sin \theta}$ to which it is assumed to be proportional) as a function of latitude, φ, from 55° N. to 55° S.; the available data did not justify going beyond these limits of latitude. Assuming the mean ocean stress to be representative of the entire latitude belt, the total torque about the axis of the Earth is given by

$$T = 2\pi R^3 \int \bar{\tau}_x \cos^2 \varphi d\varphi, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

R being the radius of the Earth. If therefore, $\bar{\tau}_x$ be plotted against latitude, on a scale where the ordinate is proportional to $2\varphi + \sin 2\varphi$, the value of T can be measured as the area subtended between the curve and the φ-axis.

IV. Distribution of Stress with Latitude and Season

The results are shown in Figure 1. The discussion of them which follows is in broad and comparative terms so that its validity is reasonably assured despite the approximations made.

In each season a delineation into four principal zones is apparent. The position of the zones shows a progression *cum sole* with a lag of about a month or even longer, in that the identifiable features such as maxima and minima and zeros of mean stress are on the whole further north in October than in April.

The curves display a striking difference between the summer and winter strengths of the zonal circulation of the northern hemisphere. In the southern hemisphere a corresponding seasonal variation is discernible in the tropical zone, though it is of an altogether smaller magnitude, while it effectively disappears in the temperate zone.

To make comparisons between zones it is most convenient to evaluate, season by season, the total torque of each, as defined by (**4**). These are presented in Table 1. The four zones will be referred to as North and South Temperate and Tropical, the separations between the zones being set for definitive purposes according to the following convention: between temperate and tropical, the latitude of zero mean stress for the season; between the two tropical zones, the latitude of minimum mean eastward stress. The poleward limits of the temperate zones were judged by extrapolation beyond 55°, but owing to the contraction of the scale at high latitudes it is evident that no major error in totals is involved on this account. It is correspondingly clear that the torque from polar regions could be appreciable only if there existed a very strong mean surface wind circulation round the pole, and there is no evidence for this.

TABLE 1

MEAN TORQUE ABOUT EARTH'S AXIS

Zone	January	April	July	October	Annual Mean	Land Fraction (F)
N. Temperate ..	—2·6	—1·6	—0·6	—1·2	—1·4	0·55
N. Tropical	+4·0	+3·5	+1·2	+2·0	+2·5	0·3
S. Tropical	+3·2	+3·6	+4·1	+3·7	+3·7	0·2
S. Temperate ..	—2·8	—3·1	—2·8	—3·2	—3·0	0·03

Unit $= 10^{26}$ g. cm.^{2}/sec.^{2}

$= 3\cdot 15 \times 10^{33}$ c.g.s. units of angular momentum per year.

Sign convention: positive for a source and negative for a sink of eastward angular momentum.

Table 1 also includes torques obtained from the annual mean curve. These do not exactly equal the mean of the four other columns, on account of the seasonal movement of the positions defining the principal zones.

The table gives quantitative expression to certain notable features of the zonal circulation pattern and its variability. The circulation is appreciably stronger in the southern than the northern hemisphere, the difference being more marked in the temperate than in the tropical zone. Even more striking is the increase in seasonal variability from south to north. The range is four- to fivefold in the N. Temperate zone, three- to fourfold in the N. Tropical, 30 per cent. in the S. Tropical, and only 15 per cent. in the S. Temperate. The error of estimation is unknown, but may be gauged from the fact that all these variations except possibly the last are of the expected form. It must also be emphasized that the figures represent essentially the *square* of the wind speed, so that the actual material circulation varies percentually by about half the above amounts, and in problems other than those considered here it may be the material circulation rather than the stress which is of physical importance. In such problems one might well adopt the premise that the mean surface circulation of the southern hemisphere shows no seasonal variation except in position.

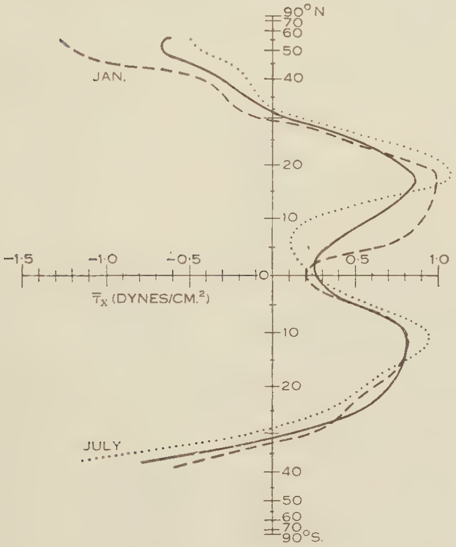


Fig. 2 (*a*)

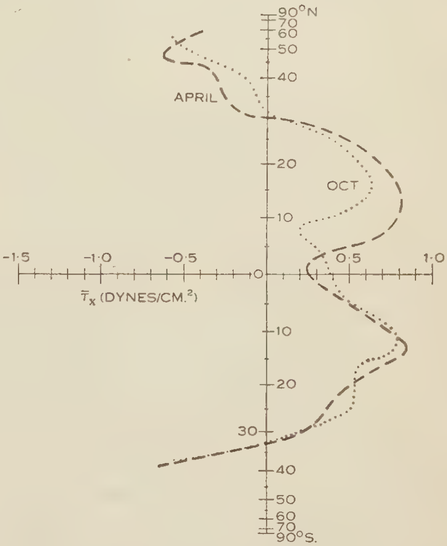


Fig. 2 (*b*)

Fig. 2 (*a*).—Mean stress distribution over Atlantic Ocean for January (dashed), July (dotted), and annual (full line).

Fig. 2 (*b*).—Mean stress distribution over Atlantic Ocean for April (dashed) and October (dotted).

Despite the relatively large percentage variation with season in the N. Tropical zone, the combined source strength of the two hemispheres varies by only about 35 per cent. through the year, from $5 \cdot 3 \times 10^{26}$ units in July to $7 \cdot 2 \times 10^{26}$ in January and April. This result has already been invoked elsewhere (Priestley 1951).

V. Distribution of Stress over the Individual Oceans

The distribution of stress with latitude for each of the oceans separately is shown in Figures 2, 3, and 4. The figures largely speak for themselves and only a very summary discussion is required.

The distributions for the Indian Ocean are quite unlike those for other oceans, in all seasons save January the north tropical easterlies disappearing

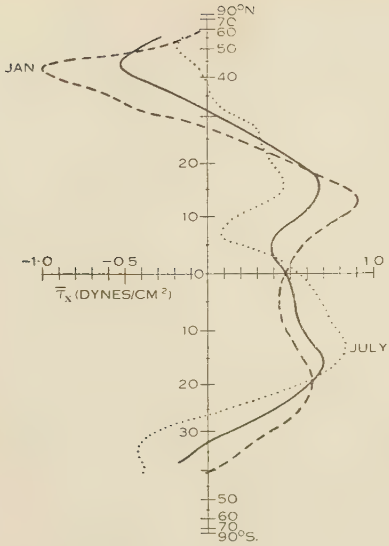


Fig. 3 (*a*)

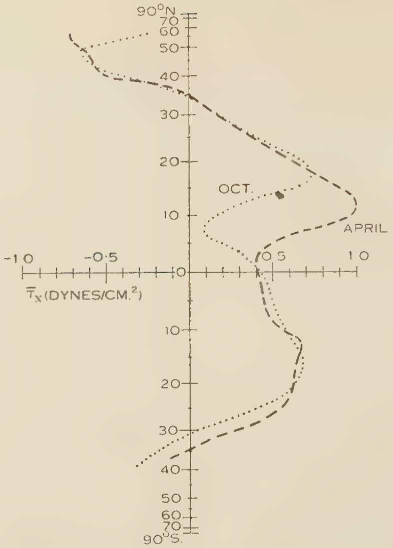


Fig. 3 (*b*)

Fig. 3 (*a*).—Mean stress distribution over Pacific Ocean for January (dashed), July (dotted), and annual (full line).

Fig. 3 (*b*).—Mean stress distribution over Pacific Ocean for April (dashed) and October (dotted).

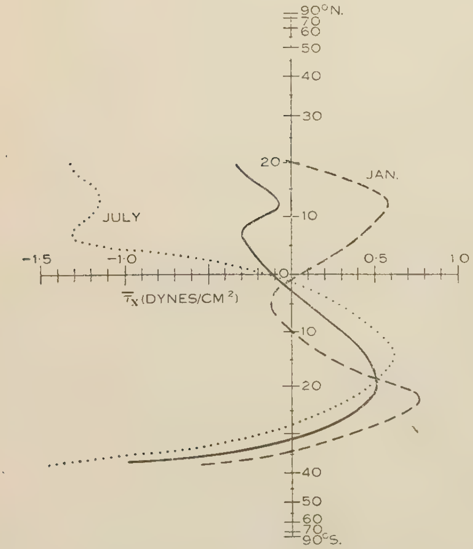


Fig. 4 (*a*)

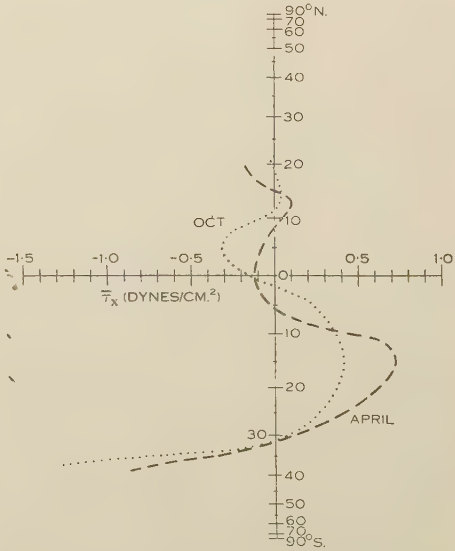


Fig. 4 (*b*)

Fig. 4 (*a*).—Mean stress distribution over Indian Ocean for January (dashed), July (dotted), and annual (full line).

Fig. 4 (*b*).—Mean stress distribution over Indian Ocean for April (dashed) and October (dotted).

effectively and being replaced by westerlies, the extreme case occurring in July. These are, of course, manifestations of the Asian monsoon system, and provide the explanation for the appearance in the July curve of Figure 1 of a minor zone of mean surface westerlies separating the two major tropical circulations. What is more surprising is that even in January, when one might have expected enhanced easterlies between the equator and 20° N., the easterlies are weaker than over the other oceans. The influence of the monsoon extends some way into the southern hemisphere, for the mean annual torque of the S. Tropical zone of the Indian Ocean is only about half that of its value elsewhere.

The Atlantic and Pacific Ocean distributions are identical in pattern, but close inspection does show some differences in degree which are not all in the sense expected. The zonal circulations of the Atlantic are if anything stronger than those of the Pacific, a difference in this sense being indicated in both hemispheres. A most noticeable feature is that the maximum of westerlies is found further northward ; in winter this reflects the more northerly situation of the Icelandic as compared with the Aleutian low pressure, but the feature is present in the summer season also. The separation between the two tropical zones is the more clearly marked in the Atlantic ; this may be only the statistical result of greater mobility, between one synoptic situation and another, of the inter-tropical convergence zone of the Pacific.

It is thought that the differences might be attributed to the fact that the Atlantic, being the narrower ocean, is more subject to land influences, and therefore to a somewhat stronger gradient of temperature with latitude. To set against such an explanation, however, is the indication in the figures that the annual variation is the more marked in the Pacific. This is shown in a more pronounced progression *cum sole* of the principal features and a very much more marked decline, in summer, of the zonal circulation of the northern hemisphere.

Fuller discussion of minor differences is hardly justified at the present stage, and the main emphasis should be on the close similarity between the two major oceans.

No discussion of the figures from the oceanographic viewpoint is offered here, but workers in this field also may find them a useful source of reference.

VI. Flow of Angular Momentum in the Atmosphere

As indicated in the Introduction, it is important to know the pattern of flow of angular momentum in the atmosphere, not only for the year as a whole but also at particular seasons. Perhaps the most fundamental issue, certainly the most controversial one at present, relates to the respective parts played by the mean meridional circulation and the quasi-horizontal disturbances (depressions and anticyclones) as agents in the poleward transport of energy, momentum, water vapour, etc. Estimates of the total transport of energy and water vapour, latitude by latitude, are available, and though these are of necessity uncertain, they are of great interest and value. From the torque values here presented it is possible, and would seem desirable, to provide corresponding estimates of the total transport of angular momentum, even though these similarly may ultimately be found to represent no more than first approximations to the true values.

For the year as a whole the total source of angular momentum to the atmosphere must equal the total sink to a very close approximation. The figures in Table 1 do not display this characteristic, the source being about 40 per cent. greater than the sink ; the discrepancy must be accounted for and an adjustment made before the flux distribution can be inferred with any pretensions to accuracy.

Part of the discrepancy may well be due to the composite effects of the land masses, i.e. to their irregular distribution over the globe, their greater effective roughness, and also (White 1949) to the torques due to pressure differences across mountain barriers. Any direct assessment of the total of these effects would be extremely crude owing to the high variability of land roughness. Nevertheless a bold assessment of the total torque for the whole of the N. Temperate zone has been made for the month of January 1946, and it is of interest to note that the result quoted by Starr (1950) does agree with the corresponding figure in Table 1, and this for the zone where the percentage of land mass is greatest. If Starr's figure be taken to imply that in the N. Temperate zone the land torque per unit area does not differ greatly from the ocean torque, balance could only be achieved by a 30 per cent. reduction in the total source strength : and since land covers only about one-quarter of the source zone this would imply a negligible torque from the tropical land masses. This interpretation is most unlikely.

Accordingly an alternative means of adjusting for the land masses was sought. Let T_0 denote the ocean torque per unit area for a given zone and F the fraction of land area in that zone. It will be assumed that the land torque per unit area is λT_0, where λ is a constant which from physical argument would be expected to be positive and greater than unity. On this basis the total torque for a zone would be given by

$$(1-F)T_0+\lambda F T_0.$$

The value of λ to give a balance between total source and sink may then be determined. This test has been applied to the annual mean figures in Table 1 but the solution found, λ about -2, was not physically acceptable : any value of λ greater than unity only increased the discrepancy between source and sink. The test, crude as it is, serves to show that balance cannot be achieved by a simple proportionate allowance for enhanced torque of the land masses. This does not, of course, deny that any correction should be made for the land effects, but it does indicate the desirability of examining whether the major part of the difference may not rather be attributed to some source of error in the ocean evaluations themselves.

Two such sources of error have already been discussed in Section II ; namely the underestimation due to a probable increase of $\varkappa$ with V, and the effect which wind variations may have on the assumed law (1). It was shown that both these sources would lead to an underestimation of the mean stress in the temperate as compared with that in the tropical zones : so that the 40 per cent. discrepancy in Table 1 is in the sense which might be expected, while its magnitude lies within the limits already assigned to the underestimation error. Part of the 40 per cent. is certainly due to the neglect of variations of density with latitude. Reference to mean isotherms readily shows that this would account for a discrepancy of order 10 per cent., again in the same sense as that shown in Table 1. (Allowance for this 10 per cent. density correction would not have altered the results of the land-factor tests.)

It would then appear reasonable, for the purpose of the approximation required, to allow balance considerations to provide the guide as to the magnitude of the relative underestimation of the stress in temperate latitudes. In order to proceed to flux evaluations, therefore, the contributions of the polar zones about 60 to 90° will be ignored, and the mean stresses for the temperate zone will be increased by 40 per cent. The pattern of northward or southward annual flux may then be inferred by starting from zero datum (taken as the poleward limit of the N. Temperate zone) and equating the flux to the cumulative net total of the source or sink strength, according to hemisphere, between the datum and the latitude in question. The results are shown in Table 3.

It is also of great interest to determine how this flux pattern changes from season to season. For the months of January and July which are times of extremes, and therefore of relatively slight change, essentially the same procedure has been followed. In these cases, however, the increase of 40 per cent. in temperate latitudes does not effect an exact balance : the out-of-balance residual has been assigned equally to fluxes into or out of the two temperate zones, as the case may be, through their poleward extremities. Such a procedure of course has no physical reality, and means that the derived flux forms may be quite wrong towards the extreme latitudes. The quantities involved are indicated in brackets opposite latitudes 60°. These figures indicate the extent to which the whole derived distributions may be in error due to this unbalanced residual.

The uncertainty due to this unbalanced residual is greatest in April, and even there its magnitude is insufficient to affect the main distinguishing features of the distribution. The evaluations for April and October have included additional allowance for the fact that the pattern of zonal momentum is in process of change at these seasons. Mean absolute angular momentum distributions in 5 degree latitude zones for winter and summer were calculated, the total mass of air in each latitude being obtained from surface pressure charts (Brunt 1939) and the angular momentum relative to the Earth from the cross sections of Hess (1948) up to 18 km. and Loewe and Radok (1950) up to 16 km. One-quarter* of the difference between winter and summer was apportioned as the gain (or loss) of angular momentum in that latitude for October and April and these amounts were added to (or subtracted from) the cumulative totals of source and sink strength, due regard being paid to sign. It has been assumed that the cross sections are representative of all longitudes ; the magnitudes are such that any error thus arising would not radically change the resulting flux pattern.

---

* This would be the appropriate factor for a sinusoidal annual variation with extremes in mid-January and mid-July.

The calculation of the differences in angular momentum between January and July, and the appropriate adjustment for "storage" during October and April, is set out in Table 2. It is apparent that the main seasonal variation

TABLE 2

CHANGE OF ATMOSPHERIC ANGULAR MOMENTUM, JANUARY MINUS JULY

Latitude Zone	Mean Pressure Diff. Jan. *minus* July (mb.)	Equivalent Diff. of Angular Momentum at Earth's Speed of Rotation ($\times 10^{30}$ g. cm.^{2} sec.$^{-1}$)	Angular Momentum Relative to the Earth ($\times 10^{30}$ g. cm.^{2} sec.$^{-1}$)			Change in Total Angular Momentum Jan. *minus* July ($\times 10^{30}$ g. cm.^{2} sec.$^{-1}$)	Assumed Rate of Gain (Loss) during Oct. (Apr.) ($\times 10^{24}$ g. cm.^{2} sec.$^{-2}$)
			January	July	Difference		
60–55 °N.	+3·0	+3	+40	+43	—3	0	0
55–50	+3·8	6	71	69	+2	+8	+1
50–45	+4·8	10	97	68	+29	39	4
45–40	+5·0	14	148	66	82	96	9
40–35	+5·2	17	228	75	153	170	16
35–30	+5·5	22	245	43	202	224	21
30–25	+5·2	24	160	—11	171	195	18
25–20	+4·5	24	124	—98	222	246	23
20–15	+3·0	18	57	—99	156	174	16
15–10	+1·2	8	4	—78	82	90	8
10– 5	0	0	—45	—59	14	14	1
5– 0	—0·2	—1	—60*	—45*	—15	—16	—1
0– 5 °S.	—2·0	—13	—53*	—25*	—28	—41	—4
5–10	—2·2	—14	—28*	0	—28	—42	—4
10–15	—5·0	—31	+3	+41	—38	—69	—6
15–20	—6·0	—35	38	112	—74	—109	—10
20–25	—6·8	—36	69	184	—115	—151	—14
25–30	—6·0	—28	86	215	—129	—157	—15
30–35	—3·8	—15	91	153	—62	—77	—7
35–40	—1·5	—5	109	99	+10	+5	0
40–45	0	0	101	79	22	22	+2
45–50	+0·5	+1	83	67	16	17	+2
50–55	+0·5	+1	59	58	1	2	0
55–60 °S.	0	0	33	51	—18	—18	—2
Total	..	.. ..	1660	1008	652		+58

* Denotes interpolated value.

resides in the angular momentum relative to the Earth and the variation due to the seasonal mass movement, as characterized by the surface pressure change, represents on the whole a much smaller term. This need not be the case, of course, for shorter periods in individual years. The greater seasonal variation of the northern than of the southern hemisphere is again well displayed in terms of the angular momentum relative to the Earth.

The last column of Table 2 completes the material from which seasonal flux evaluations can be carried out, as described above. The results are shown in Table 3. The figures in brackets opposite latitudes 60° indicate one part of the tolerance to which the whole data must be regarded as subject. Another part comes from the neglect of the actual flux into or out of the polar caps. Reference to expression (4) or to the high-latitude contraction of the ordinate scale in the figures provides assurance that this last error also will not be great enough to change the main features of the derived flux distribution.

TABLE 3

TOTAL NORTHWARD FLUX OF ATMOSPHERIC ANGULAR MOMENTUM

Latitude	Northward Flux across Latitude Circles ($\times 10^{26}$ g. cm.^{2} sec.$^{-2}$)					Total Flow per Year ($\times 10^{33}$ g. cm.^{2}/sec.)
	January	April	July	October	Annual Mean	
60 °N.	(+0·1)	(+0·8)	(+0·3)	(−0·4)		
55	+0·3	+1·1	+0·4	−0·1	+0·4	+1·3
50	+0·7	+1·4	+0·6	+0·2	+0·7	+2·2
45	+1·3	+1·9	+0·8	+0·7	+1·3	+4·1
40	+2·0	+2·3	+0·9	+1·2	+1·7	+5·4
35	+2·7	+2·5	+1·0	+1·5	+2·0	+6·4
30	+3·3	+2·4	+1·0	+1·6	+2·1	+6·7
25	+3·2	+1·9	+0·7	+1·5	+1·9	+6·0
20	+2·8	+1·2	+0·2	+1·3	+1·4	+4·5
15	+2·0	+0·3	−0·1	+0·8	+0·8	+2·5
10	+1·0	−0·6	−0·3	+0·4	+0·2	+0·5
5 °N.	+0·0	−1·5	−0·1	+0·4	−0·2	−0·8
Equator	−0·4	−1·8	−0·2	+0·1	−0·6	−1·8
5 °S.	−0·8	−2·2	−0·9	−0·4	−1·0	−3·3
10	−1·2	−2·7	−1·8	−1·1	−1·7	−5·2
15	−1·7	−3·5	−2·8	−1·9	−2·4	−7·7
20	−2·5	−4·2	−3·6	−2·8	−3·2	−10·2
25	−3·2	−4·7	−4·1	−3·5	−3·8	−12·1
30	−3·8	−4·9	−4·2	−3·9	−4·0	−12·6
35	−4·0	−4·8	−3·6	−3·7	−4·0	−12·6
40	−3·6	−4·2	−2·5	−2·9	−3·4	−10·6
45	−2·8	−3·5	−1·5	−1·8	−2·3	−7·3
50	−1·5	−2·2	−0·7	−0·8	−1·4	−4·4
55	−0·8	−1·5	−0·4	0·0	−0·7	−2·1
60 °S.	(−0·2)	(−0·8)	(−0·3)	(+0·4)		

There is in general a very considerable seasonal variation in the magnitude of the flux, particularly in the northern hemisphere where the maximum is in winter and minimum in summer. In the southern hemisphere the annual variation is less, with the maximum moving to autumn in consequence of the flux of angular momentum from the northern hemisphere during that season, coinciding with the decline of the strong northern zonal circulation. This season dominates the net annual exchange of angular momentum between the hemispheres, though there still appears a slight residual flow southwards in both January and July.

South of 30 °S. it is seen that the summer flux exceeds the winter flux in magnitude. This result, anomalous at first sight, is not inconsistent with those of Loewe (1948) who finds in this region a greater summer variability of several elements, among them temperature, humidity, cloudiness, sunshine, and precipitation. Strangely, Loewe reports no corresponding maximum variability in summer of surface wind, but his other results suggest a greater variability of upper wind, and hence a greater angular momentum flux. It may also be noted from Table 2 that the average zonal momentum is greater in summer than in winter from 35 to 55 °S.

In January the pattern between hemispheres is approximately symmetrical, but in other seasons the flux is notably stronger in the southern hemisphere. The maximum of flux occurs close to 30° latitude in both hemispheres in all seasons. The maximum of flux density, or flux per unit length, will accordingly be somewhat to poleward of this latitude.

It would be appropriate to conclude with an estimate of accuracy even though this must of necessity be subjective, taking into account the several steps involved in deriving the fluxes from the original data. There can be very little doubt that the main features of the flux distributions, with their latitude and seasonal variations, are correctly portrayed by Table 3, except at the extremities. With regard to absolute values it is thought highly probable that, for example, the peak values around 30° latitude have if anything been underestimated ; that the real value should lie in the range between the figure quoted and a value 50 per cent. higher. This would represent a very useful degree of precision, but still points the need for further research into the wind-stress relation both over the open oceans and also over land areas of varying types and covers.

VII. ACKNOWLEDGMENT

Acknowledgment is made to Mrs. J. W. Willis and other members of the Section of Meteorological Physics, C.S.I.R.O., for the labour of extraction and computation on which this paper is based.

VIII. REFERENCES

BRUNT, D. (1939).—" Physical and Dynamical Meteorology." 2nd Ed. (Cambridge Univ. Press.)
DEACON, G. E. R. (1950).—*Nature* **165** : 173.
HESS, S. L. (1948).—*J. Met.* **5** : 293.
LOEWE, F. (1948).—Commonw. Aust. Met. Bur. Bull. No. 39, p. 37.
LOEWE, F., and RADOK, U. (1950).—*J. Met.* **7** : 58.
MUNK, W. H. (1947).—*J. Mar. Res.* **6** : 203.
PRIESTLEY, C. H. B. (1951).—*Quart. J. R. Met. Soc.* **77** : 200.
STARR, V. P. (1950).—*Quart. J. R. Met. Soc.* **76** : 492.
SVERDRUP, H. U., JOHNSON, M. W., and FLEMING, R. H. (1946).—" The Oceans." (Prentice-Hall : New York.)
WHITE, R. M. (1949).—*J. Met.* **6** : 353.

THE STRUCTURE OF STEADY-STATE ANTICYCLONES

By R. W. James*

[*Manuscript received November* 9, 1950]

Summary

It is found that symmetrical anticyclones of the Australian region exhibit roughly similar pressure profiles.

The total pressure rise within a high is approximately proportional to the area covered by the disturbance, large systems in higher latitudes being more intense than small systems in lower latitudes.

These structural properties are shown to be a consequence of the requirement for gradient-wind balance within an anticyclone, which requirement is shown to limit the intensity of a high of specified size.

The theory indicates that the variation of the size of anticyclones should be restricted, and that disturbances in higher latitudes should have higher mean intensities. Both these requirements are borne out by observation.

I. Introduction

The factors determining the scale of atmospheric disturbances have been discussed from time to time notably by Jeffreys (1922) and Rossby (1938), but interest in the problem has been mainly incidental. A more explicit approach was made more recently by Priestley (1948), who discussed limits on the size of semi-permanent systems arising from the dynamical control of pressure variations due to curvature effects. The problem of what determines the intensity of systems has barely been discussed at all, but in the following it will be shown that in the case of an anticyclone, the size of the disturbance limits the intensity, or maximum pressure rise within it, if the configuration is everywhere to be in gradient-wind balance.

We shall discuss the problem here in terms of models which necessarily represent a considerable simplification of the properties of atmospheric anticyclones. The axis of the model high will be taken as vertical, and circular symmetry is assumed, although the case of elliptic symmetry is discussed in Appendix I. Winds are assumed to be everywhere horizontal, and the variation of the coriolis parameter within the model is ignored. Nevertheless, it is believed that even with such simplification the models considered represent in broad outline the structural properties of atmospheric anticyclones.

II. The Potential Energy of an Anticyclone

As shown by the writer (1950) the total potential energy, thermal and gravitational combined, of an anticyclone is

$$P=\frac{c_p}{k}\int(p-p_0)d\sigma dz, \qquad \dots\dots\dots\dots\dots\dots \quad (1)$$

---

* Section of Meteorological Physics, C.S.I.R.O., Melbourne.

where $(p-p_0)$ is the excess pressure at a given level within the vortex, p_0 being the pressure at the same height " outside " the anticyclone. c_p is the specific heat of air at constant pressure, and k is the gas constant.

We assume that at a given height the vortex has a horizontal pressure profile of the form

$$p=p_0+h(z)f(r/R), \quad \dots\dots\dots\dots\dots\dots\dots \quad (2)$$

where h is the *intensity* of the vortex, i.e. the excess of pressure at the centre, and $f(r/R)$ is the *reduced pressure profile.* By definition $f(0)=1$. R is a scale factor specifying the *size* of the vortex, and we may express distance in units or R. Thus we may define the *reduced radius*, s, by

$$s=r/R.$$

R will be defined by the condition that it is the radius at which the pressure gradient is a maximum. It is termed the *spread radius.* By definition

$$-\frac{df}{ds}\leq-\left(\frac{df}{ds}\right)_{s=1}=|f'|_{max}.$$

The reduced profile f is essentially monotonic decreasing, the pressure gradient being always directed away from the centre of the vortex, and hence f' is essentially negative. f becomes zero at the *extent* or radius at which vortical influence ceases. The extent may be finite or infinite. If the latter, f must decrease at great distance more rapidly than $1/s^2$ in order that the total potential energy should remain finite, as an inspection of equation (3) below will verify.

The potential energy of an anticyclone with the profile (2) is

$$P=\frac{2\pi c_pR^2}{k}\int_0^\infty hdz\int_0^\infty fsds. \quad \dots\dots\dots\dots\dots \quad (3)$$

We have

$$\int_0^\infty fsds=\tfrac{1}{2}s^2f\Big|_0^\infty-\tfrac{1}{2}\int_0^\infty f's^2ds.$$

Since for a vortex with finite potential energy $f\to0$ more rapidly than $1/s^2$,

$$\int_0^\infty fsds=-\tfrac{1}{2}\int_0^\infty f's^2ds,$$

and we may write

$$P=\frac{-\pi c_pR^2}{k}\int_0^\infty hdz\int_0^\infty f's^2ds. \quad \dots\dots\dots\dots \quad (3a)$$

In general the intensity may change with height, but in order that the potential energy should be finite we notice h must decrease with height at great height more rapidly than $1/z$.

Of particular importance is the *barotropic* vortex. In a frictionless atmosphere a baroclinic vortex would tend to change into a barotropic system, in which wind speed remains constant with height. We assume, therefore, that the barotropic disturbance is the typical one for a frictionless atmosphere. The condition for such a vortex is

$$h/\rho=h_0/\rho_0$$

at all heights, where ρ is the density, the suffices referring to surface conditions, density being measured outside the vortex.

The total potential energy of a barotropic vortex is thus

$$P = \frac{-\pi c_p R^2 h_0 p_{00}}{k g \rho_0} \int_0^\infty f' s^2 ds, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

p_{00} being the surface pressure outside the vortex. In future we shall write $p_{00}=p_0$, where no confusion can arise with the previous usage of p_0, as the atmospheric pressure at a given level " outside " the vortex.

III. Anticyclonic Rotation

We assume that winds are horizontal, and tangential to the stationary isobars, with no frictional forces operating. In these conditions the gradient-wind equation applies, or

$$v = \frac{\lambda \xi}{2}\left(1 - \sqrt{1 - 4v_g/\lambda\xi}\right), \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

where v is the gradient-wind speed, v_g is the geostrophic speed, ξ the radius of curvature of an isobar, and λ is the coriolis parameter, $2\omega \sin \varphi$. In what follows we shall adopt the " flat-earth " hypothesis, i.e. φ is assumed not to vary throughout the vortex, and to have the value proper to the latitude of its centre.

The equation (5) immediately shows that there is a limit on the pressure gradient, expressed by the geostrophic wind speed at a point with isobaric curvature, ξ, for if $v_g > \frac{1}{4}\lambda\xi$, the expression for v will be complex. We express this familiar result physically in the form of

Lemma I

A necessary (but not sufficient) condition that an anticyclone should be everywhere in gradient-wind balance is that the geostrophic-wind speed is everywhere less than or equal to $\frac{1}{4}\lambda\xi$.

In a circular vortex the equation (5) becomes

$$v = \frac{\lambda R s}{2}\left(1 - \sqrt{1 - 4v_g/\lambda R s}\right).$$

The geostrophic-wind speed is given by

$$v_g = \frac{1}{\lambda\rho}\,\frac{\partial p}{\partial n}.$$

For a circular vortex this may be written

$$v_g = \frac{h}{\lambda\rho R}\cdot\frac{\partial f}{\partial s}.$$

We notice that v_g is constant with height when h/ρ is constant with height, a result used in the preceding section for determining the total potential energy of a barotropic vortex.

Substituting for v_g in (5) we find

$$v = \frac{\lambda R s}{2}\left(1 - \sqrt{1 + \frac{8h}{\rho\lambda^2 R^2}\,\frac{f'}{2s}}\right). \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (6)$$

The dimensionless parameter

$$Q = 8h/\rho\lambda^2 R^2$$

has been defined elsewhere (James 1950) as the *vortex number*, the relevant density being that at the spread radius.

For a barotropic anticyclone the vortex number is invariant with height.

From (6) we see that the condition for gradient balance* everywhere is that

$$Q(-f')/2s \leq 1, \text{ or } Q \leq (2s/-f')_{min} = Q_E. \quad \ldots\ldots\ldots\ldots (7)$$

Q_E is termed the *gradient number*, and depends only on the form of the reduced pressure profile.

We can express the limitation of Lemma I in the form of

Lemma II

A necessary (but not sufficient) condition for gradient-wind balance everywhere within a circular anticyclone is that the vortex number is less than the gradient number at all levels.

In a barotropic anticyclone the vortex number is constant at all heights. In a baroclinic system Q varies with height, and the effective limitation takes place at the level of maximum Q, which is also the level of maximum wind speed. Q may equal Q_E at this level, but must be less at all other levels.

IV. THE TOTAL ENERGY OF A CIRCULAR BAROTROPIC ANTICYCLONE

The expression for the total kinetic energy of a circular vortex is

$$K = \int \tfrac{1}{2}\rho v^2 d\sigma dz.$$

Using the expression (6) for v we find

$$K = \frac{\pi\lambda^2 R^4}{4}\int_0^\infty \rho dz \int_0^\infty s^3(1-\sqrt{1+Qf'/2s})^2 ds. \quad \ldots\ldots\ldots\ldots (8)$$

For a barotropic vortex

$$K = \frac{\pi\lambda^2 R^4 p_0}{4g}\int_0^\infty s^3(1-\sqrt{1+Qf'/2s})^2 ds. \quad \ldots\ldots\ldots\ldots (9)$$

The total energy of a barotropic anticyclone,

$$\begin{aligned} E &= P + K \\ &= \frac{\pi\lambda^2 R^4 p_0}{4g}\int_0^\infty \left[-\frac{c_p}{k}\,\frac{Qf'}{2s},+(1-\sqrt{1+Qf'/2s})^2\right] s^3 ds. \quad \ldots\ldots (10) \end{aligned}$$

If the reduced pressure profile is specified, we see from the form of (9) that E is an increasing function of Q. But the condition of gradient-wind balance limits Q, such that

$$Q \leq Q_E,$$

---

* The strict conditions

$$\frac{Q\rho_0(-f')}{2\rho s} \leq 1,$$

but ρ/ρ_0 is always within a few per cent. of unity. The maximum horizontal variation of density within an isothermal vortex is only 1 per cent. per 10 mb. of intensity.

where Q_E depends only on the reduced profile. Hence the maximum possible energy of an anticyclone with prescribed spread and reduced profile is

$$E_M = \frac{\pi\lambda^2 R^4 p_0}{4g}\int_0^\infty \left[-\frac{c_p}{k}\frac{Q_E f'}{2s} + (1-\sqrt{1+Q_E f'/2s})^2\right] s^3 ds. \quad \ldots\ldots (\mathbf{11})$$

Alternatively, if the total energy of an anticyclone is prescribed, along with the reduced profile, the expression (**11**) gives a minimum possible scale (spread) for the anticyclone.

The expression (**10**) can be written

$$E = \frac{16\pi p_0 h_0^2}{\lambda^2 g \rho_0^2}\int_0^\infty \left[-\frac{c_p' f}{2kQs} + \frac{1}{Q^2}(1-\sqrt{1+Qf'/2s})^2\right] s ds. \quad \ldots\ldots\ldots (\mathbf{12})$$

The expression under the integral sign is a decreasing function of Q. Hence for a given total energy, the maximum possible intensity is given by the relation $Q = Q_E$, f being specified, or, alternatively, a minimum possible total energy is associated with a given intensity.

The total energy of an anticyclone depends on the reduced profile shape. It may be that there is an " optimal " profile, which permits the greatest possible energy concentration in a given area, of extent radius, A, say.

From (**10**) we have

$$E = \frac{\pi\lambda^2 p_0}{16g}\int_0^{A^4} \left[\frac{c_p}{k}\psi + (1-\sqrt{1-\psi})^2\right] dr^4, \quad \ldots\ldots\ldots\ldots\ldots (\mathbf{13})$$

where $\psi = -Qf'/2s$.

For a given A the greatest possible value of E is obtained when $\psi = 1$ for all r. The reduced pressure profile satisfying this condition, with $f(0) = 1$, is

$$f = 1 - s^2/Q, \quad s \leq A/R, \qquad f = 0, \quad s > A/R.$$

We have $f' = -2s/Q$, so that from the condition $\psi = 1$ we have $Q = Q_E = 1$.

Hence from (**13**) the total energy of the " optimal " anticyclone is

$$E_M = \frac{\pi\lambda^2 R^4 p}{16g} \cdot \frac{c_p + k}{k}. \quad \ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{14})$$

The velocity profile of the optimal anticyclone is, simply,

$$v = \lambda r/2, \quad r < R, \qquad v = 0, \quad r > R,$$

i.e. a solid rotation with angular speed equal to one-half of the coriolis parameter. The anticyclonic vorticity relative to the Earth is $-\lambda$, i.e. equal in magnitude, but opposite in sign to the local vertical component of the Earth's vorticity. Hence

Lemma III

The greatest possible energy for a circular, barotropic anticyclone of specified size is developed by a solid rotation with zero absolute vorticity.

In terms of the intensity, the expression (**14**) for the energy of this optimal rotation is given by

$$E_M = \frac{4\pi h_0^2 p_0}{\rho_0^2 \lambda^2 g} \cdot \frac{c_p + k}{k}. \quad \ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{15})$$

A proportion $c_p/(c_p+k)$ of this energy is in potential form, while the remainder, $k/(c_p+k)$, or 22 per cent., is in the form of kinetic energy. It is not difficult to show that a solid rotation gives a greater proportion of kinetic energy than any other circular wind distribution.

In the case of the *normal profile* (James 1950)

$$f=\partial e^{-s2/2} \qquad \text{for all } s,$$

the equation (**10**) may be written

$$E_M=\frac{\pi\lambda^2R^4p_0}{4g}\int_0^\infty\left[\frac{c_pQ}{2k}\partial e^{-s2/2}\left(1-\sqrt{1-\frac{Q}{2}\partial e^{\ s2/2}}\right)^2\right]s^3ds.$$

The maximum energy is developed in a barotropic normal vortex when

$$Q=Q_E=2.$$

The potential energy developed in this case is

$$P_m=\pi c_p\lambda^2R^4p_0/2gk,$$

and the kinetic energy is only about 2 per cent. of this, so that

$$P_m\approx E_m.$$

Comparing this expression with the corresponding one for a solid rotation, expression (**15**), we notice that the total energy in both cases is proportional to λ^2R^4, or, alternatively, to

$$h_0{}^2/\lambda^2.$$

The total energy for a barotropic anticyclone, with $Q=Q_E$, can, in fact, be represented by an equation of the form

$$E=Mh_0{}^2/\lambda^2, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \textbf{(15a)}$$

or, alternatively,

$$E=NR^4\lambda^2, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \textbf{(14a)}$$

where M and N depend only on the profile shape. If $Q<Q_E$ we must specify Q as well as the profile shape. If, in addition, the anticyclone is non-barotropic we must multiply M and N by the ratio

$$Z/Z_B,$$

where Z is the *equivalent height* (James 1950) of the vortex, defined by

$$h_0Z=\int_0^\infty hdz$$

and Z_B is the equivalent height of a barotropic vortex

$$Z_B=p_0/\rho_0g\approx 8 \text{ km}.$$

It will be seen subsequently that there is evidence for supposing that Q is only slightly, if at all, lower than Q_E in steady-state anticyclones, and that Q_E does not vary greatly between different disturbances. Hence it is reasonable to suppose that the variation in total energy as between different disturbances in the same latitude is predominantly due to the variations in intensity (or spread).

An anticyclone of intensity 2 mb. is about as " weak " as the system could be, and still be recognized as a closed circulation. On the other hand, intensities of 40 mb. are rarely exceeded in the Earth's atmosphere. An intensity range of 20 to 1 corresponds to an energy range of 400 to 1.

If the range of anticyclonic energies is of the order 400 to 1, the range of spread in anticyclones will be of the order $400^{\frac{1}{4}}$ to 1, i.e. only 4·5 to 1, from (**14a**).

Petterssen (1940) quoted the mean length of the longest axis of subtropical anticyclones in summer, 5900 km., with a standard deviation of 1840 km. Thus 69 per cent. of anticyclones have extents (taken as half the length of the longest axis) ranging between 2000 and 3900 km., representing an energy range of 13 to 1.

Thus we reach the conclusion that all anticyclones are roughly of the same size, because the size of an anticyclone depends very weakly on the total energy, and the range of total anticyclonic energy is not excessively great, at any rate for " measurable " disturbances.

The relation (**15a**) shows that the energy of an anticyclone of given intensity depends on latitude, through λ, low latitude highs having a greater total energy. With a given intensity an anticyclone centred in latitude 30° will have four times the energy it would have centred on the pole.

This latitude effect can be eliminated by defining a *potential intensity* H, by the relation

$$H = h/\sin\,\varphi,$$

H being the intensity which a high would achieve if removed from its present latitude to the pole without change in the total energy. The equation (**15a**) can then be written

$$E = M'H^2,$$

where $M' = M/4\omega^2$.

Similarly we can define a *potential spread*

$$L = R\,(\sin\,\varphi)^{\frac{1}{2}},$$

and write (**14a**)

$$E = N'L^2,$$

where $N' = 4\omega^2 N$.

We may write the vortex number

$$Q = 8h/\rho\lambda^2 R^2 = 2H/\rho\omega^2 L^2.$$

The vortex number is thus invariant to change of latitude if the anticyclonic energy remains constant.

V. The Role of Friction

In an attempt to apply the foregoing ideas to the atmospheric anticyclone, certain limitations must be borne in mind. We have treated the case of gradient-wind balance. In the actual atmosphere, winds are never perfectly gradient, even in the steady state. Frictional drag prevents the achievement of perfect gradient balance, departures from the ideal wind being greatest near the surface.

Consider a solid anticyclonic rotation to be set up in the real atmosphere. At the outer confines of the vortex there is a discontinuity of wind between the vortex and the surrounding, stagnant air. Under these circumstances lateral frictional coupling will tend to transfer momentum from the vortex to the ambient air. Pressure profiles would adjust themselves to the new wind field, and we might anticipate a final steady state with a wind profile different from that of the initial solid rotation. The effect of the frictional coupling is to impose a constraint on the form of the steady-state wind profile. The nature of the modification is shown graphically in Figure 1. The maximum wind speed is reduced from V_0 to V, and the spread radius from R_0 to R. Although the spread of a disturbance is reduced by frictional influence, the extent is increased, the wind profile extending beyond R_0. The general role of friction is to spread out the area of influence in this sense.

Our quantitative knowledge of frictional forces in the free atmosphere is at the present time in too primitive a stage to be made the basis of a theory of the shape and intensity of naturally occurring anticyclones. It suffices to remark that the influence of friction is least where winds are lightest, and any approach to the "optimal" form of a solid rotation should be closest near the centre. Many writers have accepted the solid rotation as typical of anticyclonic motion in the inner parts of a vortex.

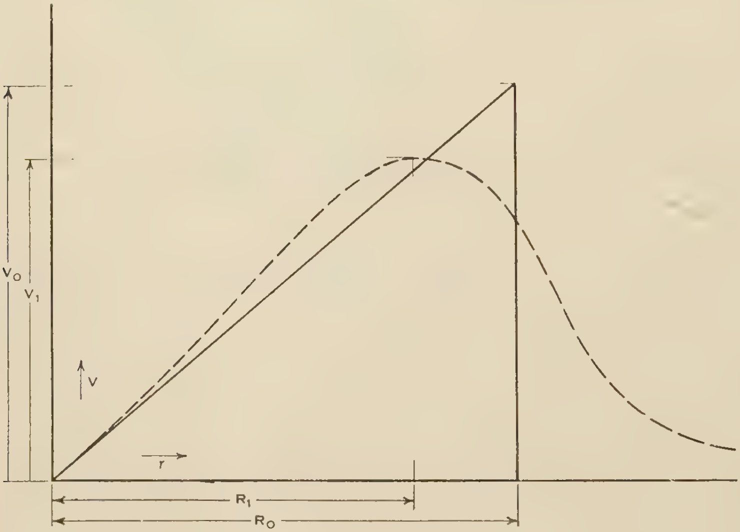


Fig. 1.—Velocity profiles for a solid rotation and a wind field modified by friction.

VI. The Gradient-Wind Limitation

The type of barotropic anticyclone we have been considering is specified by three parameters, the spread, the intensity, and the profile shape. If the gradient-wind condition is an effective limiting factor, the number of independent variables is reduced to two, and if, further, the principle of maximum possible energy concentration is valid, one parameter alone suffices to determine the steady-state vortex.

We must now enquire whether there is, in fact, such an association between the three "independent" parameters of an anticyclone. This can only be done by determining empirically whether the relations between the parameters suggested by theory are in fact satisfied in actual anticyclones. That task is reserved to the section immediately following. We offer here certain " plausibility " considerations.

An anticyclone is a region of overnormal total energy. This energy can not derive from processes going on within the vortex area, for a high may be regarded as deficient in energy sources. It is true that the warm type anticyclone is, over land at least, normally associated with clear skies and net insolation warming of the lower layers, particularly in summer. However, it is generally agreed that there is a net radiation loss from the higher levels of an anticyclone, and this high level sink may outweigh the effect of the low level source. That it normally does so is evident from the consideration that anticyclones tend to be more intense in winter than in summer. The converse would be the case if surface heating were the significant factor, and in any event, intense surface heating in the summer is usually associated with heat lows.

The excess energy of an anticyclone must therefore be derived by importation from cyclonic regions which, while themselves deficient in energy, are the seat of energy sources as, e.g. of latent heat release. This passive view of the anticyclonic function was probably first propounded by the late Sir Napier Shaw, who regarded highs as the " dumping ground " for the " debris " thrown out of cyclonic areas.

On this view it would be natural to assume that the extent of anticyclonic development is determined by the vigour of cyclogenesis taking place elsewhere. If the cyclonic activity is great, and the " dumping ground " of small geographical extent, an anticyclone might soon absorb all the energy it can assimilate. Energy assimilation could take place in three ways, by a vertical extension of the vortex, i.e. by an approach towards barotropy, by a change in the horizontal profile in the direction of a solid rotation, or by an increase in the size of the system.

Cold " shallow " highs show a tendency to develop into warm " deep " ones. This may be interpreted as an attempt at energy assimilation by an approach towards barotropy. At the same time a high often " spreads ", another factor tending towards increased assimilation.

Finally, improved assimilation might be achieved by a change in the profile shape, although frictional influences may inhibit the realization of a perfect solid rotation.

There are thus reasons for supposing that the vortex number, Q, approximates closely to the limiting gradient number, Q_E, for some anticyclones at least.

VII. The Vortex Number of Anticyclones

The profile shape is defined by the condition $f(0)=1$, $f(S)=0$, where S is the extent radius, and $|f'(1)|$ is a maximum. These conditions, along with that which prescribes that f is a monotonic function already restricts the shape of permissible profiles considerably. The conditions, in fact, amount to a two-parametric representation of f. If, in addition we specify the numerical value of $-f'(1)$, we restrict possible profiles to a much narrower class of function.

We are, in fact, employing a three parametric representation of f.

With the accuracy of meteorological measurement what it is, an improved fit to a set of observations obtained by the use of four parameters rather than three is rarely statistically significant. Hence for all practical purposes, the spread, the extent, and the profile slope at the spread radius determine *the* profile shape, i.e. determine it as accurately as observation will permit.

Instead of the value of $-f'(1)$ as a parameter determining the profile shape, it will be found more convenient to use the inverse

$$\Gamma=-2/f'(1)=-(2s/f')_{s=1}\ ;$$

Γ might be termed the *inverse flexure* (James 1950).

Now we have

$$Q_E=(-2s/f')_{min}.$$

But Γ is a particular value of $(-2s/f')$, and hence cannot be less than the minimum value, or

$$\Gamma\geq Q_E.$$

Since $|f'(1)|$ is by definition the maximum value of $|f'|$, we have

$$|f'(1)|<\text{average}\,|f'|=1/S,$$

where S is the ratio of the extent to the spread, whence

$$Q_E<2S$$

is a generous upper limit to Q_E, and hence, also, to Q.

For atmospheric anticyclones S may be of the order, 3, and seldom much greater than this. If we take as an upper limit, $S=5$, we have 10 as the upper limit to Q and Q_E. The lower limit is, of course, zero.

Let us now consider the values of Q_E and Γ for a few profile shapes.

The "optimal" solid rotation is represented by

$$f=1-s^2,\quad s\leq1,\qquad f=0,\quad s\geq1,$$

and it is easy to show that $Q_E=\Gamma=1$.

The *sinoidal profile* is defined by

$$f=\tfrac{1}{2}(1+\cos\pi s/2),\quad s\leq2,\quad f=0,\quad s\geq2.$$

The extent of this profile is twice its spread, and we find

$$Q_E=1\cdot63,\quad\Gamma=2\cdot55.$$

Attention need not be confined to profiles of finite extent. A suitable form for an infinite profile is that representing the *normal vortex* (James 1950)

$$f=e^{-s^2/2}.$$

For this type of structure $Q_E=2$, $\Gamma=3\cdot3$.

Another form of profile of infinite extent is

$$f=\left(1+\frac{n-1}{n+1}s^n\right)^{-1},$$

representing the *power vortex* (James 1950). For this profile we have

$$Q_E=\frac{8n}{n^2-4}\left(\frac{n-2}{n+2}\cdot\frac{n+1}{n-1}\right)^{2/n},$$

$$\Gamma=8n/(n^2-1).$$

With $n=2$, $Q_E=3$, $\Gamma=5\cdot3$; for $n=4$, $Q_E=2\cdot0$, $\Gamma=2\cdot1$. When n is very large both Q_E and Γ decrease to zero as $8/n$.

Γ can be estimated from synoptic data much more easily than Q_E.

It would be of interest to compute the values of Q and Γ for persistent anticyclones. The migrating anticyclones of the Australian region are cells in the subtropical high belt, and show marked persistence. They are therefore admirably suited to the present enquiry. In so far as they are migratory, the gradient-wind balance cannot be perfect, but in the case of slow moving systems at least this distortion cannot be of great moment. Most of the systems studied were distinctly elliptical in shape; hence for the spread radius we must substitute the semi-minor spread, b (see Appendix I).

Table 1 gives estimates of the parameters of 20 roughly symmetrical anticyclones recorded on the Australian surface weather maps for 0900 hr. during the period October-December 1948. Data given are: an identification number for each anticyclone, date of occurrence and central latitude, and estimates of the intensity, and semi-minor spread. The vortex number and Γ derived from these figures are given in the last two columns.

TABLE 1

VORTEX NUMBERS FOR ANTICYCLONES

No.	Date (1948)	Lat.	h (mb.)	b (miles)	Q	Γ
1	Oct. 3	34	5	300	2·0	1·2
2	„ 9	25	7	380	3·1	3·5
3	„ 11	29	8·5	330	3·8	1·9
4	„ 14	34	13	500	1·9	2·6
5	„ 15	33	7·5	420	1·6	2·5
6	„ 19	33	5	360	1·1	1·0
7	„ 22	37	17	500	2·1	3·4
8	„ 23	32	15	700	1·2	2·8
9	Nov. 1	33	10	600	1·1	1·5
10	„ 8	30	6	500	1·1	1·2
11	„ 13	40	8	400	1·4	1·5
12	„ 17	34	9	500	1·3	1·9
13	„ 18	34	11	550	1·3	1·5
14	„ 22	40	19	600	1·6	2·8
15	„ 23	38	9·5	400	1·8	1·9
16	Dec. 3	32	9	300	4·0	3·0
17	„ 6	35	11	350	3·1	3·1
18	„ 20	38	10	420	1·7	2·6
19	„ 28	36	11	400	3·3	2·6
20	„ 30	40	18	420	2·3	2·5

Accurate values of these two last quantities are difficult to obtain, for they both depend on h and b, which can only be estimated approximately. Individual values, particularly of Q, which depends on b^2, may thus be largely in error.

The mean vortex number of the 20 anticyclones is 1·9, with a standard deviation of 0·9, while the mean Γ is 2·3, with a standard deviation of 0·7.

In five cases out of 20 Q is greater than Γ. This is not possible for an anticyclone in gradient-wind balance, and the discrepancy may be put down to errors of estimation.

To test for subjective factors in the evaluation of the quantities involved, independent estimates of Γ for anticyclones 9–19 inclusive were made by another analyst. The analysts agreed to within 10 per cent. on $\overline{\Gamma}$, and the coefficient of correlation between the two sets of estimates was 0·73, suggesting that about half the variation in Γ was due to errors in estimation.

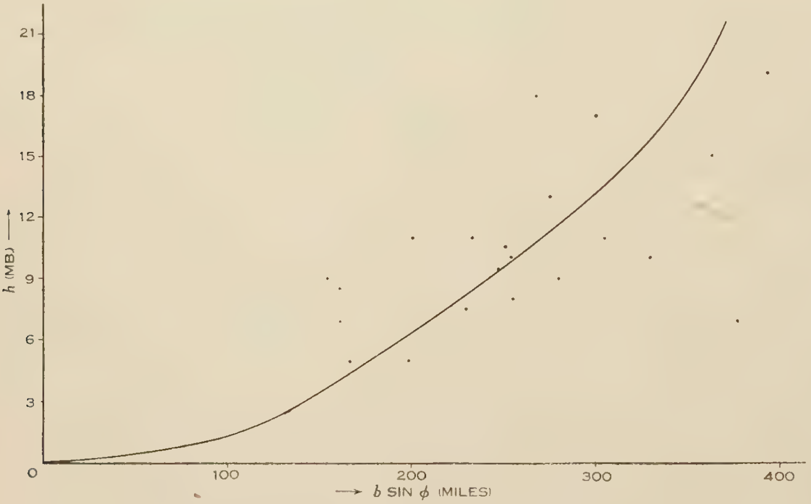


Fig. 2.—Intensity and spread of anticyclones.

Figure 2 is a scatter diagram of intensity against b sin φ. A definite correlation will be noted, high values of h being associated with high values of b sin φ. The regression parabola (minimizing normal deviations) is indicated by the full line

$$h=\frac{\overline{Q}}{2}\rho\omega^2 b^2 \sin^2\varphi.$$

$\overline{Q}$, the regression constant is identified with the mean Q. If Q is constant h varies with b sin φ. The coefficient of correlation between h and $b^2 \sin^2 \varphi$ is 0·73, suggesting that one-half of the observed variation in h is due to variation in b sin φ, the remaining variance being due to variations in Q from vortex to vortex, and to experimental error. The variance due to experimental error is large (as shown by the difference in independent estimates of Γ) and hence there would seem to be little residual variance to be accounted for by real variation in Q. Hence the data are quite consistent with the hypothesis that Q and Γ vary little, if at all, in different anticyclones.

We are thus led to the conclusion that the intensity of an anticyclone is specified within fairly narrow limits once the spread is given. We have

$$h \approx \tfrac{1}{4}\rho\lambda^2 R^2,$$

or, numerically, in latitude 30°,

$$h \text{ (mb.)} = 4\cdot 5 \times 10^{-5} R^2 \text{ (miles}^2\text{), or}$$

$$h = 1\cdot 7 \times 10^{-5} R^2 \text{ (km.}^2\text{)}.$$

An interesting feature of the relation between intensity and spread is its dependence on latitude. In latitude 60° the intensity for a given spread is three times that given for latitude 30°, and at the pole, four times. Thus we should expect anticyclones to be of generally higher intensity in the higher latitudes, as can be shown to be the case statistically.

The observed mean value of the vortex number, $\bar{Q}=1\cdot 9$ is definitely higher than the gradient number of a solid rotation, or a sinoidal vortex, and is approximately the same as Q_E for a normal vortex, and an $n=4$ power vortex. $\bar{\Gamma}=2\cdot 3$ is not greatly in excess of the value of this parameter for an $n=4$ power vortex, but is significantly less than Γ for a normal profile. Thus we might take as the " best " analytic representation of the mean pressure profile for anticyclones

$$f=(1+0\cdot 6s^4)^{-1}.$$

The approximate equality of $\bar{Q}$ and $\bar{\Gamma}$ establishes that

$$\bar{Q}=Q_E, \text{ very nearly.}$$

Hence the experimental data establish that the gradient-wind condition is, in fact, an effective constraint on the structure of atmospheric highs.

VIII. CONCLUSIONS

We have considered the properties of circular, barotropic anticyclones which may be represented by three parameters, the intensity, the spread, or scale, and Γ, specifying the profile shape.

From the relatively small spread in the determined values of Γ, much of which can be attributed to experimental error, we infer that the profile shape does not vary widely for different highs. The mean anticyclone profile can be represented to a degree of approximation by

$$f=(1+0\cdot 6s^4)^{-1}.$$

Given the approximate constancy of Γ, it is found that there is a limitation upon the values which h and R may assume independently if the condition of gradient balance is to be satisfied everywhere within the vortex. This limitation is expressed by the relation.

$$Q \leq Q_E,$$

where Q_E depends only on the profile shape.

If Q were much less than Q_E, the control implied by gradient balance would not be an effective limiting factor on anticyclonic structure, and h and R could be independent of one another.

If, however, the intensity and size of anticyclones tend to be such that Q is close to Q_E, h and R are no longer independent, the relation between them being

$$8h/\rho\lambda^2 R^2 \approx Q_E \approx 1\cdot 9. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (16)$$

An anticyclone of given profile shape is then completely determined by one parameter only, h, or R, or the total energy.

That gradient-wind control is effective in determining the structure of some anticyclones at least is evident from the fact that in the 20 cases studied, the mean vortex number is only slightly less than the mean Γ. Since a condition of gradient balance is $\Gamma \geq Q_E$, we must have

$$\tilde{Q} \approx \bar{Q}_E,$$

i.e. gradient balance effectively limits the intensity of a high of specified size and profile.

Since the range of values of h in individual anticyclones is of the order 20 : 1, the range of R is only of the order $4\frac{1}{2}$: 1, from (**16**). Hence highs of widely varying intensity should be of roughly comparable size. That atmospheric highs are relatively restricted as to size is, then, further evidence of the operation of the gradient-wind constraint.

Finally, the indication of the theory is that the mean intensity of highs should increase with latitude. This conclusion, too, is borne out by observation.

We conclude that the requirements of gradient-wind balance do effectively limit the structure of atmospheric highs, and impose a relation of the form (**16**) on the intensity and spread.

IX. REFERENCES

JAMES, R. W. (1950).—*Quart. J. R. Met. Soc.* **76** : 255.

JEFFREYS, H. (1922).—*Quart. J. R. Met. Soc.* **48** : 29.

PETTERSSEN, S. (1940).—"Weather Analysis and Forecasting." p. 150. (McGraw-Hill Book Co. : New York.)

PRIESTLEY, C. H. B. (1948).—*Quart. J. R. Met. Soc.* **74** : 67.

ROSSBY, C. G. (1938).—*Pap. Phys. Oceanogr. Met. Camb. Mass.* **7** : 1.

APPENDIX I

Elliptical Anticyclones

Anticyclones in nature are more often ellipse-shaped than circular, and it is important to determine what bearing this fact may have on the condition of gradient balance.

The form of pressure profile consistent with a solid rotation, at angular speed, $\lambda/2$ up to the spread radius R, is

$$p=p_0+h(1-r^2/R^2), \quad r \leq R, \qquad p=p_0, \quad r \geq R.$$

Let us now distort this pressure profile so that the isobars take on an elliptical shape,

$$x^2/a^2+y^2/b^2=\text{constant}.$$

The distorted pressure profile is

$$p=p_0+h(1-x^2/a^2+y^2/b^2), \quad x^2/a^2+y^2/b^2 \leq 1,$$

$$p=p_0, \quad x^2/a^2+y^2/b^2 \geq 1.$$

The *spread ellipse* is defined by

$$x^2/a^2+y^2/b^2=1.$$

Instead of a single spread we have $(a>b)$ a semi-minor spread, b, and a semi-major spread, a.

The gradient wind in such a system is

$$v=\tfrac{1}{2}\lambda\xi\left[1-\sqrt{1-\frac{8h}{\rho\lambda^2\xi}(x^2/a^4+y^2/b^4)^{\frac{1}{2}}}\right],$$

where ξ, the radius of curvature of an isobar as (x, y), is given by

$$\xi=a^2b^2(x^2/a^4+y^2/b^4)^{3/2}/(x^2/a^2+y^2/b^2).$$

Eliminating ξ between these two equations we find

$$v=\frac{4h(x^2/a^4+y^2/b^4)^{\frac{1}{2}}/\rho\lambda}{\left(1+\sqrt{1-\frac{8h}{\rho\lambda^2b^2}\cdot\frac{x^2/a^2+y^2/b^2}{x^2/a^2+y^2/b^2+(a^2/b^2-1)y^2/b^2}}\right)}.$$

The necessary condition for gradient-wind balance everywhere is

$$\frac{8h}{\rho\lambda^2b^2}\cdot\frac{x^2/a^2+y^2/b^2}{x^2/a^2+y^2/b^2+(a^2/b^2-1)y^2/b^2}\leq 1,$$

that is $\dfrac{8h}{\rho\lambda^2b^2}\leq 1.$

The condition

$$Q=8h/\rho\lambda^2b^2\leq 1$$

is identical with the condition of gradient balance in a solid rotation, with the semi-minor spread replacing the spread of a solid rotation.

Although the proposition is only proved for an elliptically distorted solid rotation, the result holds true for any profile. The gradient number for $f[(x^2/a^2+y^2/b^2)^{\frac{1}{2}}]$ is the same as for the corresponding circular profile, $f(s)$.

CONJUGATED COMPOUNDS

III. SPECTRA AND IONIZATION IN POLYENES AND POLYMETHINES BY A FINITE DIFFERENCE METHOD*

By N. S. Bayliss† and J. C. Rivière‡

[*Manuscript received February* 27, 1951]

Summary

Periodic potential energies were calculated for π-electrons along the molecular axes of butadiene and three polymethines NH_2—$(CH=CH)_n$—$CH=NH_2^+$, using $Z'_C=1\cdot0$ and $Z'_N=2\cdot0$ or $1\cdot5$ as the respective effective nuclear charges of the carbon and nitrogen nuclei. The Schroedinger equation was solved by finite differences, and with two restricted and self-consistent parameters, good agreement was obtained with observed ionization energies, band frequencies, and band intensities. The Brillouin zone effect is discussed, and also the screening in conjugated compounds. The use of an iteration procedure for the solution of eigenvalue problems by finite differences is described in Appendix I.

I. Introduction

The essential feature of the " free electron " method for discussing the spectral behaviour of conjugated molecules is the assumption of a simple and explicit form for the potential energy field of the π-electrons of the conjugated system, followed by a direct solution of the Schroedinger equation containing the assumed potential energy. In the earlier papers by Bayliss (1948), Kuhn (1948*a*), and Simpson (1948), the potential energy of the π-electrons was given the extremely simple form of an electron gas at uniform potential in a one-dimensional space bounded by infinite potential walls. In the preceding paper of this series, Bayliss (1950) showed that the π-electron potential energy along the axis of a conjugated system should be periodically oscillating rather than uniform, and also that the original " box " with infinite potential walls should be modified to give the potential energy the shape of a " well " with walls of finite height corresponding to the existence of the finite experimental ionization energies of π-electrons. However, in order to obtain convenient analytical solutions of the Schroedinger equation, he used non-periodic functions that, on the average, approximated the calculated periodic potential energy functions.

In the present paper we deal with the periodic form of the π-electron potential energy along the molecular axis of a linear conjugated system, and describe the application of a finite difference method to obtain numerical solutions of the Schroedinger equation for the cases of butadiene and several linear polymethine ions. The assumptions that are involved and the method of calculation will first be described in relation to the butadiene molecule.

---

* From a thesis presented by J.C.R. for the degree of M.Sc. in the University of Western Australia. The material was presented in part to the Symposium on Molecular Structure and Spectroscopy, Ohio State University, U.S.A., in June 1950.

† Department of Chemistry, University of Western Australia, Nedlands, Western Australia.

‡ Present address: Division of Industrial Chemistry, C.S.I.R.O., Melbourne.

II. Butadiene

(a) *The Potential Energy*

As in the previous paper (Bayliss 1950), the one-dimensional potential energy curve of a π-electron along the molecular axis is considered in two parts according as the electron is outside or inside the electronic cloud of the molecule. Outside the electronic cloud (i.e. outside the points YY′ in Fig. 1), the electron is regarded as being in the field of the residual charge $+e$ which is assumed to be located at the centre of the molecule. Inside the electronic cloud (i.e. inside the points YY′), the electron is regarded as in the field of the four carbon nuclei and of the remaining (π, σ, and K) electrons. We assume that the effect of the remaining electrons can be regarded as one of screening the carbon nuclei, which are therefore each assigned an effective (screened) nuclear charge $Z'e$. The hydrogen nuclei are neglected, i.e. are assumed to be completely screened.

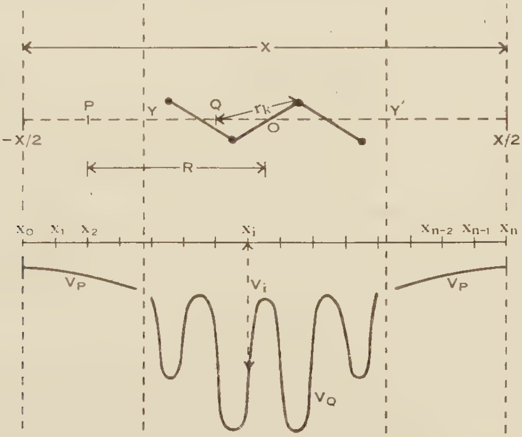


Fig. 1.—Method of calculating π-electron potential energies along the axis of a linear conjugated compound. Butadiene is shown drawn to scale in the upper half of the figure, and the portions of the potential energy curve corresponding to equations (1) and (2) on the same scale in the lower half. The significance of the points Y and Y′ can be readily understood.

The molecular axis (see Fig. 1) is chosen as the x-axis, and the origin O is placed at the centre of the molecule. At all points P *outside* YY′, the potential energy V_P is given by

$$V_P=-e^2/R, \quad \ldots\ldots\ldots\ldots \quad (1)$$

where R is the distance PO. At all points Q *inside* YY′, the potential energy V_Q is given by

$$V_Q=\Sigma_k(-Z'e^2/r_k), \quad \ldots\ldots\ldots\ldots \quad (2)$$

where r_k is the distance from Q to the kth nucleus, and the summation is extended over the four nuclei, which are thus assumed to be equally screened regardless of the position of Q in the range Y to Y′. As the π-electron under consideration moves outward from O along the axis, the potential energy V_Q must join smoothly to V_P at the points Y and Y′ where the electron moves from inside to outside the electronic cloud. The effect of this in drawing the potential energy curve is that V_Q and V_P cannot be referred to the same formal zero of potential energy, since if they were, V_Q would always lie below V_P at Y and Y′. The zero of potential energy for our problem is the zero of V_P, corresponding to ionization as $R \rightarrow \infty$. The condition of continuity at the points YY′ means that the formal zero of V_Q must be adjusted to make possible the smooth junction at Y and Y′. The amount of this adjustment depends on the positions chosen for Y and Y′, which thus become parameters to be discussed later.*

Assuming for the moment that the positions of Y and Y′ are known, the potential energy V is readily calculated at any point along the molecular axis. Such a curve is shown in Figure 1, where in addition it has been assumed that all the C–C distances are 1·39 Å and that all the C–C–C bond angles are 120°. It will be obvious from the nature of the calculations that this slight departure from the experimental values of Schomaker and Pauling (1939) can have no significant effect on the results.

(*b*) *Solution of the Schroedinger Equation by Finite Differences*

The Schroedinger equation is

$$d^2\psi/dx^2+(8\pi^2m/h^2)(E-V)\psi=0. \quad \ldots\ldots\ldots\ldots (3)$$

To solve it by finite differences we confine the problem to the finite range of length X between $x=x_0=-X/2$ and $x=x_n=+X/2$. This range is divided into n equal parts (the " net " in relaxation terminology) by the $(n+1)$ equidistant points $x_0 \ldots . x_i \ldots . x_n$. The calculated potential energy at each point x_i is denoted by V_i. The usual boundary condition of the Schroedinger equation ($\psi=0$ at $x=\infty$) must now become the finite boundary condition that $\psi=0$ at the boundaries $x=x_0$ and $x=x_n$. The distance X, though finite, may be made as large as desired, the one condition being that it should be large enough so that the calculated ψ's should become sensibly zero well within the chosen boundaries. It will be seen later that this condition is not difficult to fulfil.

Equation (3) is converted to a dimensionless form by substituting $z=x/X$, and putting

$$\lambda=(8\pi^2mX^2E)/h^2\,; \quad k=(8\pi^2mX^2)/h^2, \quad \ldots\ldots\ldots\ldots (4)$$

thus giving

$$(d^2\psi/dz^2)+\lambda\psi-kV\psi=0. \quad \ldots\ldots\ldots\ldots (5)$$

The variable z ranges from $+\frac{1}{2}$ to $-\frac{1}{2}$, and corresponding to the x_i there are $(n+1)$ points z_i separated by equal intervals $a=1/n$. Let V_i and y_i denote the values of V and of ψ at z_i; the differential $(d^2\psi/dz^2)$ at z_i becomes in finite difference form

$$(d^2\psi/dz^2)_i=(y_{i+1}-2y_i+y_{i-1})/a^2. \quad \ldots\ldots\ldots\ldots (6)$$

---

* An analogous problem in classical electrostatics would be the potential energy due to a point charge $+q$ that is surrounded by a Faraday cage bearing the charge $-(q-1)$.

Thus for each point z_i one obtains for equation (5)

$$y_{i+1}-(2-a^2\lambda+a^2kV_i)y_i+y_{i-1}=0. \quad \ldots\ldots\ldots\ldots\ldots \quad (7)$$

This set of $(n+1)$ equations can be solved for the y_i only when λ (containing the energy E) has certain eigenvalues $\lambda_1, \lambda_2 \ldots, \lambda_N \ldots,$ corresponding to the energy levels $E_1 \ldots, E_N \ldots,$ of which E_1 may be taken to be the lowest, and so on. The problem is thus to find the λ_N, and for each of these to find the eigenfunction ψ_N in the form of a set of numerical values $y_0, y_1 \ldots y_i \ldots y_n$ at the $(n+1)$ points of the net. Southwell's (1946) relaxation method was used in the early stages of the work ; it was then replaced by an iteration method deduced from it which was considerably simpler in dealing with the one-dimensional problem considered here, although it is not obviously adaptable to the multidimensional cases that can be handled by relaxation.

(c) *Results of the Calculations*

We will leave the description of the computational method to Appendix I and deal here with the results of the calculations on butadiene after first discussing the selection of values for the parameters and similar quantities.

Several test calculations were made to determine how close the boundaries of the problem could be brought to the molecule without significantly affecting the results. The boundaries that were finally chosen were very conservative as shown by the way in which the eigenfunctions in Figure 6 approach zero well inside the points $x=x_0$ and $x=x_n$.

The number $(n+1)$ of the points in the net is also subject to arbitrary choice although it is not in the strict sense a parameter. The more points in the range, the more accurately will the equations (7) give the eigenvalues and eigenfunctions corresponding to a given potential energy curve. In the present case we performed trial calculations until increasing the number of points had no further appreciable effect on the solution. With the chosen boundaries, 29 points were sufficient for butadiene, corresponding in general to $(4n+13)$ points for a conjugated system with n nuclei.

One of the most important parameters is the value assigned to Z'. Using simpler, approximate potential energy curves, Bayliss (1949, 1950) showed that $Z'\approx 1$ for carbon atoms in linear polyenes, and justified this conclusion by a rather arbitrary extension of Slater's (1930) screening rules. It will be seen below that of several trial values the choice of $Z'\approx 1$ again gives the best agreement with experiment.

The other important parameter is the location of the points YY′ where the " inside " and " outside " potential energy curves join ; this in turn is related to the amount by which the formal energy zero of equation (2) for V_Q is displaced from the energy zero of the problem (i.e. the zero of equation (1)). In the previous paper (Bayliss 1950) it was shown that there was no great latitude in the choice of YY′. On physical grounds they must lie outside the terminal carbon nuclei, and yet not too far outside but rather at some distance that might be regarded as the average location of the " edge " of the π-electron cloud. In the previous paper (Bayliss 1950) it was assumed that YY′ are beyond the end nuclei by about half the projected length on the axis of a C–C bond (0·6 Å).

In the present work, these junction points have not necessarily coincided with points on the net, and eventually it was found convenient to draw the V_Q and V_P curves as if the points YY′ corresponded to a potential energy equal to that of the highest maximum within the V_Q curve. This located Y and Y′ at about 0·23 Å beyond the end carbons.

(*d*) *Calculations on Butadiene*

The trial calculations on butadiene will now be described in some detail. Four models (or sets of parameters) were used, as follows:

I. This first model, calculated in the early stages of the investigation, actually was a modification of the original model with infinite ionization energy as used by Bayliss (1948), Kuhn (1948*a*), and Simpson (1948).

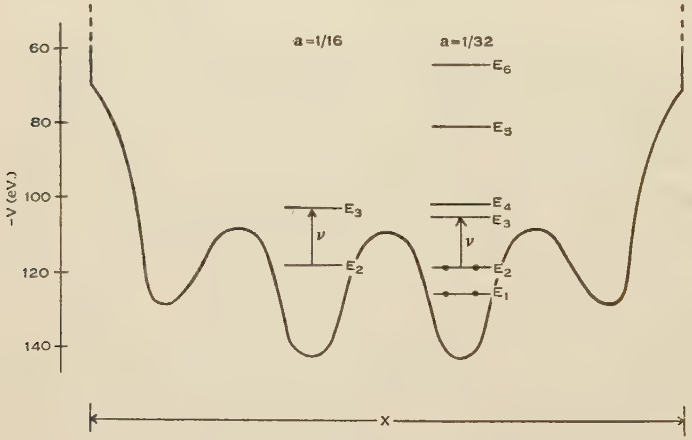


Fig. 2.—Preliminary model for butadiene, showing energy levels and the potential energy curve calculated from equation (2) with $Z'_C=2$.

The "outside" portion of the potential curve V_P was omitted, and vertical potential energy walls were drawn at the points YY′ terminating the V_Q curve (see Fig. 2). The boundary condition was that $\psi=0$ at Y and Y′. We assumed $Z'=2$ as a compromise between the Slater (1930) value of $Z'=3\cdot2$ for an isolated carbon atom and the lower value of $Z'=1$ suggested in the previous paper (Bayliss 1950). Two nets were used, with $a=1/16$ and 1/32.

II. The V_P and V_Q curves were drawn according to equations (1) and (2), the zero of the V_Q curve being displaced so that the two curves joined at the points YY′ located 0·53 Å beyond the end carbon nuclei. The boundaries of the problem were placed at 3·72 Å outside the end carbons (see Fig. 3). In this model we used the Slater value of $Z'=3\cdot2$, and calculations were made for nets with $a=1/18$, 1/36, and 1/72.

III. This model retained $Z'=3\cdot2$, placed the points YY′ at 0·23 Å outside the end carbon nuclei, and also brought the boundaries of the problem nearer to the end carbons at a distance of 1·9 Å. The net had $a=1/24$ (see Fig. 4).

IV. The final choice was to put $Z'=1$ as in the previous paper (Bayliss 1950) and to place YY′ as in model III. The boundaries were placed at 2·5 Å beyond the end carbons, and the net had $a=1/28$ (see Fig. 5).

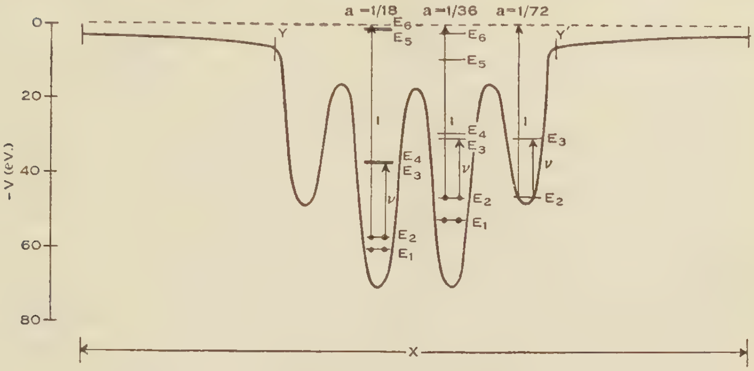


Fig. 3.—Model II for butadiene. The potential energy curve has been drawn in two parts, according to equations (1) and (2), joining at the points YY′ situated 0·53 Å beyond the end carbon nuclei. $Z'_C=3·2$ was used in the calculations, and the problem was bounded at points 3·72 Å outside the end carbon atoms.

The energy levels calculated for models I-IV are shown in Table 1, and are also drawn in Figures 2–5. The four π-electrons in the ground state of butadiene are assigned to the two lowest levels E_1 and E_2. The energy (E_3-E_2) is that

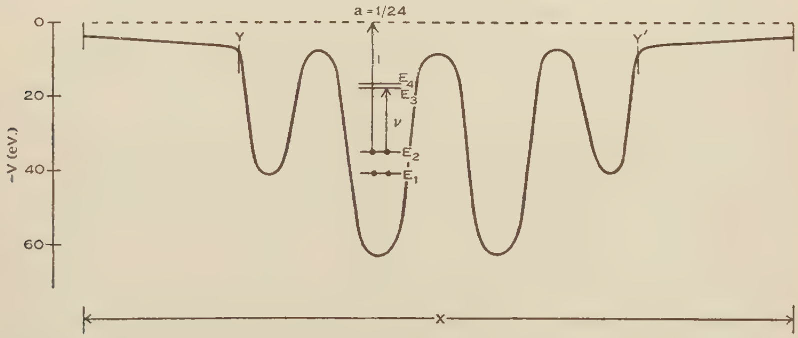


Fig. 4.—Model III for butadiene, modified from the preceding model by shifting the boundaries of the problem and the positions of the points YY′ to 1·9 Å and 0·23 Å, respectively, from the end carbon nuclei.

of the lowest optical transition, and gives a ν (calc.) to be compared with the experimental ν (exp.) for the maximum of the main absorption band. The energy of E_2 (which except in the case of model I is measured from an energy zero corresponding to ionization) gives the calculated ionization energy I (calc.) to be compared with I (exp.). In each case the eigenfunctions were also calculated (see example in Fig. 6), and from ψ_2 and ψ_3 of model IV the oscillator strength f(calc.) of the $(2 \rightarrow 3)$ transition was obtained by the method described previously (Bayliss 1948, 1950) but using numerical integration in the present instance.

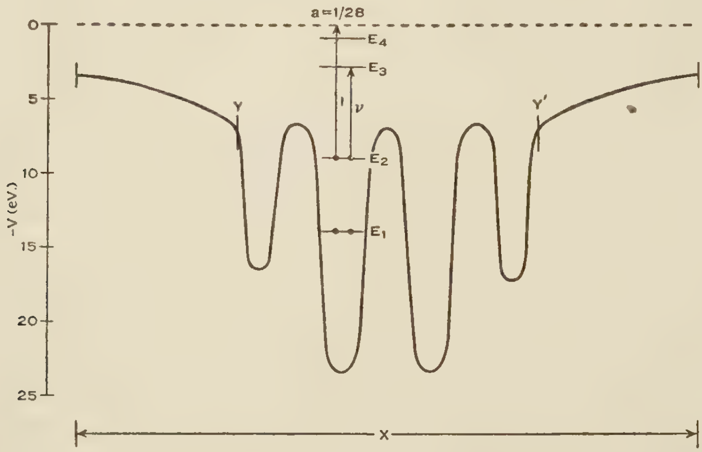


Fig. 5.—Final model for butadiene. Points YY′ are in the same positions as in Model III, but $Z'_C=1$ was used in equation (**2**), and the boundaries were placed 2·5 Å beyond the end carbons.

Table 1 shows the effect of changing the size of the net in models I and II; clearly there is little change on going from $a=1/36$ to $a=1/72$. This combined with the fact that we found that closer boundaries could be used in the final model prompted us to take $a=1/28$ in that case. The closer choice of boundary in model IV was conservative, since the calculated eigenfunctions became sensibly zero well inside them.

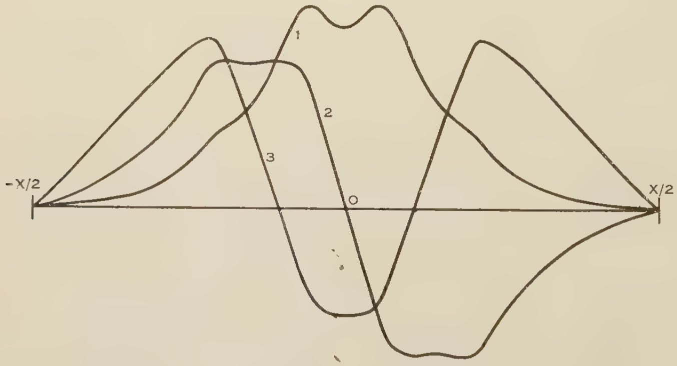


Fig. 6.—Eigenfunctions corresponding to the energy levels shown in Figure 5.

As mentioned previously, the two parameters are Z' and the junction points YY′ (or in other words the displacement of the zero of the V_Q curve). Altering Z' changes both ν(calc.) and I(calc.), the former more than the latter;

altering YY′ or the energy zero of V_Q has most effect on I(calc.). The values of both ν(calc.) and I(calc.) in models I-III are obviously far too great, those of ν(calc.) in particular suggesting that the whole energy scale should be reduced by a factor of about 3. This was additional evidence supporting the choice in model IV of $Z'=1$, a value which had already been indicated by previous work (Bayliss 1950). The agreement between theory and experiment in model IV is very satisfactory.

TABLE 1

CALCULATIONS ON FOUR MODELS FOR BUTADIENE

	Model I		Model II			Model III	Model IV
	$a=1/16$	$a=1/32$	$a=1/18$	$a=1/36$	$a=1/72$	$a=1/24$	$a=1/28$
Energy levels (volts)$^{(a)}$							
E_1		126·04	60·48	52·30		40·74	13·84
E_2	118·43	118·90	57·27	46·85	46·37	35·14	8·91
E_3	103·88	105·86	37·26	30·56	30·46	17·84	3·28
E_4		102·03	37·20	29·19		16·41	1·09
E_5		81·54	1·36	9·37			
E_6		64·76	1·26	2·20			
ν(calc.) (cm.$^{-1}$)	117380	105200	161400	131400	128350	139600	45400
ν(exp.) (cm.$^{-1}$)$^{(b)}$ ..		48000			48000	48000	48000
I(calc.) (volts)			57·27	46·85	46·37	35·14	8·91
I(exp.) (volts)$^{(b)}$ ..					9·02	9·02	9·02
f(calc.)							0·57
f(exp.)$^{(b)}$							0·53

(a) The energies are *negative*, and are measured downwards from a zero which has no physical significance in model I, but which corresponds to ionization in models II-IV.

(b) From data quoted by Bayliss (1948, 1950).

III. THE POLYMETHINE IONS $NH_2—(CH=CH)_n—CH=NH_2^+$

In order to construct the potential energy curves of polymethine ions, we transferred unmodified from butadiene (model IV) those parameters that are appropriate. The same C–C distances and bond angles were used, and for simplicity the N–C distance was assumed to be the same as C–C. The " outside " V_P curve was constructed as before, except that since the substances are themselves positive ions, equation (1) becomes

$$V_P = -2e^2/R. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

The junction points YY′ were 0·23 Å beyond the end nitrogens as in model IV of butadiene, and the boundaries were taken about 2 Å beyond the end nitrogens. The effective nuclear charge of carbon was again taken as $Z'_C=1$, and to calculate

V_Q it only remains to choose the corresponding value Z'_N for the nitrogen nuclei. Taking the polymethine with $n=1$ as an example, and regarding the π, σ, and K electrons as in successively inner shells (Bayliss 1949, 1950) to which the Slater screening constants can be applied as in the case of a single atom, any given π electron is screened from the nuclei by the other 5 π-electrons, by the 15 σ electrons (neglecting the hydrogen nuclei each with one σ electron), and by 10 K electrons. With the screening constants of 0·35, 0·85, and 1·00 respectively

TABLE 2

CALCULATIONS ON THE POLYMETHINES $NH_2—(CH=CH)_n—CH=NH_2^+$

	$n=1$ (6 π-electrons)				$n=2$ (8 π-electrons)		$n=3$ (10 π-electrons)
	$Z'_N=1\cdot5$		$Z'_N=2$		$Z'_N=1\cdot5$	$Z'_N=2$	$Z'_N=1\cdot5$
	$a=1/30$	$a=1/60$	$a=1/30$	$a=1/60$	$a=1/40$	$a=1/40$	$a=1/48$
Energy levels (volts)[a]							
E_1	17·09	17·06	21·41	20·94	17·01	22·30	14·17
E_2	15·81	15·79	21·14	20·61	15·91	22·30	13·06
E_3	12·35	12·35	14·99	14·73	14·14	13·95	12·06
E_4	7·23	7·19	10·10	9·79	11·70	10·97	10·89
E_5					8·20	7·65	8·73
E_6					4·19	3·26	5·81
E_7					0·34		2·72
ν(calc.) (cm.$^{-1}$) ..	41300	41600	39500	39850	28200	26800	23600
ν(exp.) (cm.$^{-1}$)[b] ..		35300		35300	26100	26100	20600
f(calc.)		0·80		0·57	1·07	0·74	1·29
f(exp.)[b]		0·95		0·95	1·12	1·12	1·37

(a) The energies are *negative*, measured from a zero corresponding to ionization.

(b) Experimental data from Simpson (1948), corrected for bathochromic effect of dimethyl substitution.

(Slater 1930), and taking $Z'_C=1\cdot19$ as calculated by Bayliss (1949, 1950), one obtains $Z'_N=1\cdot96$. (The Slater (1930) value for a $2s$ or $2p$ electron in the isolated N atom is 3·9.)

On the other hand, this extension of the Slater rules to a molecule may be rather arbitrary, particularly in the case of a molecule with nuclei of different atomic numbers. One would expect the higher charge on the terminal nitrogens to cause an electronic drift towards them, with consequently greater screening and lower Z'_N. We have therefore carried out calculations with two assumed values of Z'_N, namely $Z'_N=2$ as suggested by the previous paragraph, and $Z'_N=1\cdot5$ to take care empirically of the electron drift.

Table 2 gives the energy levels, and also the calculated and experimental values of ν and *f* for the polymethine ions with $n=1, 2, 3$. Figures 7 and 8 show the potential energy curves and the energy levels for the first ion ($n=1$) for the two cases $Z'_N=1\cdot5$ and $Z'_N=2$ respectively. For this ion we also performed calculations with the two nets $a=1/30$ and $1/60$; it is obvious that the coarser net is adequate. Since the polymethine ions are known only in the solid state and in solution, there are no experimental data relating to the ionization energy *I*.

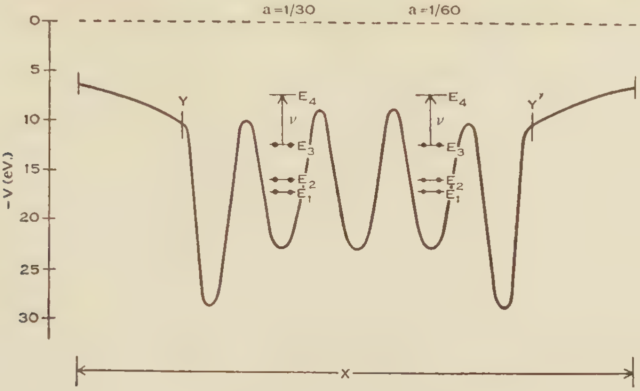


Fig. 7.—Potential energy curve and energy levels for the polymethine ion NH_2—CH=CH—CH=NH_2^+, calculated from equations (**2**) and (**8**) with $Z'_C=1$ and $Z'_N=1\cdot5$. The points YY′ and the boundary points are 0·23 Å and 2·5 Å respectively from the nitrogen atoms at the ends of the molecule.

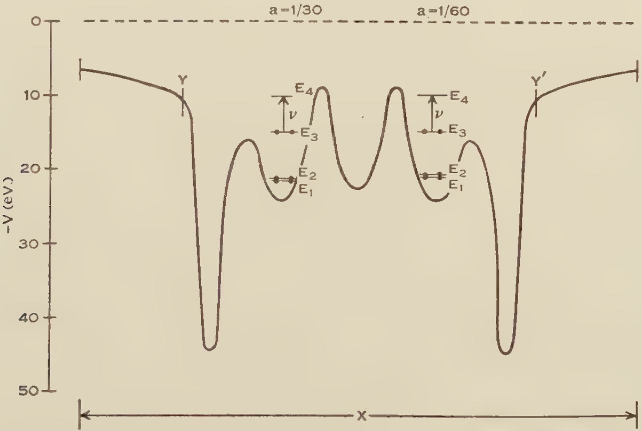


Fig. 8.—Potential energy curve and energy levels for the polymethine ion NH_2—CH=CH—CH=NH_2^+, calculated for the same values of the parameters as in Figure 7, with the exception that $Z'_N=2$.

In the cases of the ions with $n=1$ and 2, there is not much change in ν(calc.) whether one takes $Z'_N=1\cdot5$ or 2, the latter value agreeing rather better with experiment. The choice of $Z'_N=1\cdot5$, however, makes f(calc.) much closer to the experimental figure. We adopted $Z'_N=1\cdot5$ for the third ion ($n=3$) and found good agreement between theory and experiment. It should be noted that with $Z'_N=2$ the lowest energy levels tend to be grouped in a close pair, a point that is referred to later.

It is a feature of the absorption spectra of the higher members of the polymethine series (Simpson 1948) that weaker absorption bands appear to the high frequency side of the main, intense absorption band. The ion with $n=2$ has a weaker band at 44,000 cm.$^{-1}$* in addition to the intense band at 26,100 cm.$^{-1}$*. The ion has 8 π-electrons, and the intense band is identified with the transition $E_4 \rightarrow E_5$. A linear symmetrical molecule is subject to the selection rule that $\triangle N$ is odd, and furthermore the transitions with $\triangle N=3$ are much weaker than those with $\triangle N=1$ (Bayliss 1948, 1950). One is therefore tempted to identify the weak 44,000 cm.$^{-1}$ band with one of the transitions $E_2 \rightarrow E_5$, $E_3 \rightarrow E_6$, or $E_4 \rightarrow E_7$, all the energy levels being available. The least energetic of these, $E_2 \rightarrow E_5$, gives ν(calc.)$=62,200$ cm.$^{-1}$, a value which is qualitatively in the right direction but quantitatively is too large. No experimental value is available for the intensity of the 44,000 cm.$^{-1}$ band other than the statement that it is weak (Simpson 1948). Similarly the two weak high frequency bands for the ion with $n=3$ (Simpson 1948) could be accounted for qualitatively, but not quantitatively, in terms of transitions involving one of the two excited levels that have been calculated as lying between the top filled level (E_5) and ionization.

IV. Discussion

There are certain features of the calculations in the preceding section that restrict their absolute quantitative significance, for example the simplification involved in considering only one dimension. Finite difference methods could in principle be extended to three-dimensional calculations,† but the labour of computation would become forbidding. However, there are grounds for believing that, for linear molecules particularly, a one-dimensional approximation is probably very reasonable. In the simpler but related case of a free electron gas enclosed in a space in which one dimension is considerably greater than the other two (e.g. a cylinder or a rectangular prism with one long axis), it is easy to show that the energy of the type of transition with which we are concerned here is dominated by the energies of the wave functions along the long axis. The factors bearing on the choice of the molecular axis as the representative linear dimension have been discussed previously (Bayliss 1950).

It is also difficult to estimate the effect of using one-electron wave functions and energy levels, and of neglecting spin-orbit interaction. A better approximation might be obtained by using suitable linear combinations of the one-electron ψ's, but the advantages are not nearly as great here as in the LCAO molecular orbital method, since in the present case the one-electron wave functions already have the correct number of nodes (namely 0, 1, 2.... in successive energy levels), a stage that is reached in the MO method only by taking suitable LCAO's. Indeed the successive models for butadiene used by Bayliss (1948, 1950) and in the present paper show that the energy levels, the wave functions, and particularly the calculated transition probabilities are remarkably insensitive to changes in the form of the potential energy function provided it has a suitable " depth ". This suggests that the possession of the proper nodal structure is one of the most fundamental properties of the wave functions, a point that has also been emphasized by Herzfeld (1947), Walsh (1948), and Platt (1949).

---

* Both values corrected for the bathochromic effect of dimethyl groups as estimated by Simpson (1948).

† One of us (J.C.R.) has treated benzene as a two-dimensional problem using relaxation.

Platt (1949) has pointed out that with the inclusion of spin-orbit interaction, the first excited energy level in the free electron method as applied to linear molecules is really split into two levels, a singlet and a triplet. The observed intense absorption band is of course due to the singlet-singlet transition ; the location of the excited triplet state is unknown in the case of the compounds considered here, except that it must lie below the excited singlet state. The calculated ν's should therefore really be compared with the unknown mean of the singlet and triplet transitions rather than with the singlet transition itself.

Although the method inevitably employs two parameters which to some extent are subject to arbitrary choice, it is evident from the calculations of the preceding section that this choice is rather restricted. Furthermore this limited and consistent choice of parameters has led to a satisfactory reproduction of the data in both the series of polyenes (ν, I, and f) and polymethines (ν and f).* It was a defect of the simple free electron model (Bayliss 1948 ; Kuhn 1948*a* ; Simpson 1948) with uniform potential energy that the polyene and polymethine series required different relationships between the one length parameter and the actual length of the conjugated system. Kuhn (1948*b*, 1949) explained this difference in behaviour by assuming that there was a difference in the periodicity of the potential energy between the two series. In the polymethines the periodicity is that of the atomic nuclei in the conjugated chain ; in the polyenes Kuhn assumed a superimposed periodicity due to the C=C—C units owing to the alternation of bond length which is presumed to be absent in the polymethines. The effect of this longer periodicity of potential energy in the polyenes groups the energy levels into smaller Brillouin zones than in the polymethines, so that in the former series the main transition is from the highest level of one zone to the lowest level of the next, while in the latter series the transition is in the middle of a zone. Figures 3, 4, 7, and 8 show that there is in fact a grouping of the levels into zones ; but that it is not necessarily connected with an alternation of bond lengths. In fact the zones are related to the periodicity of potential energy minima of unequal depth, a point that is well illustrated by comparing Figures 7 and 8. We incline to the opinion that Kuhn has rather exaggerated the alternation of bond length in the polyenes (see Kuhn 1949, Fig. 11), and that the parameter he has employed to account for this effect in the polyenes should be regarded as a parameter and not necessarily physically significant. Lennard-Jones (1937) has shown by the molecular orbital method that there is less alternation in C=C—C bond distances as the polyene molecule becomes longer, yet it is for the longer polyenes (e.g. carotene) that Kuhn's parameter has been most successful.

---

* The same choice of parameters has given good agreement with experiment in calculations on conjugated dialdehydes—unpublished work by J.C.R.

It is of interest that the calculations confirm the low values of the effective nuclear charges Z'_C and Z'_N that have been suggested previously by Bayliss (1949, 1950). Other workers have often used the Slater (1930) atomic value of $Z'_C=3\cdot 2$ in calculations on π-electrons in conjugated compounds. However, this value was obtained by assuming the $2s$ and $2p$ electrons to be equivalent in the isolated atom, and there is little reason to expect the atomic value to hold in conjugated compounds, where the π- and σ-electrons are certainly not equivalent. The π-electrons are most loosely bound and subject to greater screening than the σ, leading one to expect $Z'_C<3\cdot 2$ for the former. On the other hand it is possibly fortuitous that the application of Slater's screening constants to the successive π, σ, and K shells gives a value for Z'_C that is in such good agreement with the model parameter. Actually the assumption in equation (**2**), that the same Z'_C is used for all the nuclei along the chain irrespective of the position of the point Q, is an over-simplification in the opposite direction from the usual one of neglecting all but " nearest neighbours ".*

However, the concept of π-electrons forming their own shell outside the σ-bonded framework and under the influence of quite distant nuclei is a qualitatively valuable one.

V. ACKNOWLEDGMENTS

This work was carried out while one of us (J.C.R.) was in receipt of a Junior Research Fellowship from the Commonwealth Research Grant. We are deeply indebted to Mr. F. S. Shaw of the Aeronautical Research Laboratories, Department of Supply, Melbourne, for providing instruction in relaxation methods, and to Mr. L. P. Coombes, the Chief Superintendent, for making temporary accommodation available in the Laboratories.

VI. REFERENCES

BAYLISS, N. S. (1948).—*J. Chem. Phys.* **16** : 287.

BAYLISS, N. S. (1949).—*J. Chem. Phys.* **17** : 1353.

BAYLISS, N. S. (1950).—*Aust. J. Sci. Res.* A **3** : 109.

HERZFELD, K. F. (1947).—*Chem. Rev.* **41** : 233.

KUHN, H. (1948*a*).—*Helv. Chim. Acta* **31** : 1441.

KUHN, H. (1948*b*).—*J. Chem. Phys.* **16** : 840.

KUHN, H. (1949).—*Z. Elektrochem.* **53** : 165 ; *J. Chem. Phys.* **17** : 1198.

LENNARD-JONES, J. E. (1937).—*Proc. Roy. Soc.* A **158** : 280.

PLATT, J. R. (1949).—*J. Chem. Phys.* **17** : 484.

SCHOMAKER, V., and PAULING, L. (1939).—*J. Amer. Chem. Soc.* **61** : 1769.

---

* An attempt by the authors to apply equation (**2**) to extreme cases such as the graphite crystal showed that equal Z'_C over the infinite plane leads to infinite energies.

SIMPSON, W. T. (1948).—*J. Chem. Phys.* **16** : 1124.

SLATER, J. C. (1930).—*Phys. Rev.* **36** : 58.

SOUTHWELL, R. V. (1946).—" Relaxation Methods in Theoretical Physics." (Oxford Univ. Press.)

WALSH, A. D. (1948).—*Quart. Rev.* **2** : 73.

APPENDIX I

*Solution of Finite Difference Equations by an Iteration Method**

We begin with the set of $(n+1)$ equations

$$y_{i+1}-(2-a^2\lambda+a^2kV_i)y_i+y_{i-1}=0. \quad \ldots\ldots\ldots\ldots \quad (7)$$

Firstly it is to be noted that the boundary condition is that $\psi=0$ at the boundaries of the problem. Thus

$$y_0=y_n=0. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (9)$$

Next assume a trial value λ' for an eigenvalue of λ, and also an arbitrary value y'_1 for y_1 (see below for notes on this procedure). A value y'_2 is then calculated for y_2 by the equation $(i=1)$

$$y'_2-(2-a^2\lambda'+a^2kV_1)y'_1+y_0=0, \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{10})$$

remembering that $y_0=0$ by definition, and that V_1 is known. Having obtained y'_2, a value y'_3 of y_3 is obtained from the next equation $(i=2)$

$$y'_3-(2-a^2\lambda'+a^2kV_2)y'_2+y'_1=0. \quad \ldots\ldots\ldots\ldots \quad (\mathbf{11})$$

Proceeding in this way one comes to a final equation which must be written

$$y_n-(2-a^2\lambda'+a^2kV_{n-1})y'_{n-1}+y'_{n-2}=R'. \quad \ldots\ldots\ldots\ldots \quad (\mathbf{12})$$

All the y''s are known in this equation since $y_n=0$ by definition, and since y'_{n-1} and y'_{n-2} have already been determined from the preceding equations. It follows that R' (the " residual " in relaxation terminology) can be zero only if λ' is actually one of the eigenvalues of λ. In general, $R'\neq 0$ at the first attempt.

A new trial eigenvalue λ'' is then taken, and the procedure repeated to find a new residual R''. In practice only a few repetitions are needed to find a trial λ that makes R sensibly zero. This λ is then one of the eigenvalues of the equations (**7**), and the y_i calculated from the equations (**10–12**) give the corresponding (un-normalized) eigenfunction ψ.

Although this procedure may seem on description to involve a great deal of repetitive calculation, it was found in practice to be quick and convenient. The following points should be noted :

(*a*) It is not always obvious at the outset just to *which* eigenvalue λ_N the successive trial values of λ' are converging. The answer is given by the number of nodes in the corresponding eigenfunction ψ_N, since there are no internal nodes if $N=1$, one if $N=2$, and so on.

(*b*) Experience in the method, and foreknowledge of the order of magnitude of the energies to be expected, are of material help in the initial choice of λ'. After two trials of a λ' and λ'', comparison of the magnitude and sign of the residuals R' and R'' makes it possible to converge fairly quickly to a λ that makes $R\approx 0$.

---

* This method was developed by J. C. Rivière.

(*c*) The arbitrariness in choosing y'_1 in equation (**10**) is only apparent, since the eigenfunction is thus far un-normalized. One of the ordinates in an un-normalized eigenfunction (in this case y_1) can always be chosen arbitrarily.

(*d*) The calculations can be shortened by making use of elements of symmetry in the potential energy function and therefore in the eigenfunctions. In the present case of linear molecules with a centre of symmetry, the calculations need be carried over only half the length of the molecule by using either one of the following conditions. If y_c represents the ordinate of ψ at the centre of symmetry then (i) $y_c=0$ if ψ has a node at the centre (antisymmetric eigenfunctions), (ii) $y_{c+1}=y_{c-1}$ if ψ has no node at the centre (symmetric eigenfunctions).

THE DIELECTRIC PROPERTIES OF A SERIES OF *dl-n*-EICOSANOLS

By R. J. Meakins* and H. K. Welsh*

[*Manuscript received January* 30, 1951]

Summary

Initial investigations of the dielectric properties of normal long-chain secondary alcohols were made with the symmetrical compounds because of their ready availability. The properties of long-chain polar compounds are known to vary considerably, however, with the position of the polar group in the carbon chain and it therefore seemed desirable to make dielectric measurements with some unsymmetrical alcohols. The present paper describes such an investigation of a series of *dl-n*-eicosanols having the hydroxyl groups in the 2-, 4-, 6-, 8-, and 10-positions, respectively.

The results for the melted forms show that slight asymmetry, as in the 10-compound, has little effect on the dielectric properties, but with the more unsymmetrical 4-, 6-, and 8-compounds a considerable enhancement of the dielectric absorption is observed. This is accompanied by increased dispersion of the dielectric constant which, for the 8-compound, reaches a low frequency value of 15. Both ε'' and ε' decrease during storage at room temperature.

Anomalous results are obtained with *dl-n*-eicosan-2-ol, probably due to the molecules in the crystal lattice being alternately reversed, end-to-end.

As with the symmetrical secondary alcohols, the recrystallized forms give comparatively little dielectric absorption.

I. Introduction

During research on the dielectric properties of organic long-chain compounds, it was recently found that certain crystalline forms of secondary alcohols give abnormally high dielectric absorption over a wide range of audio and radio frequencies (Meakins and Sack 1949, 1951). The results could be explained on the assumption that the secondary alcohol crystals contain hydrogen-bonded chains of hydroxyl groups capable of reversing their directions by rotation of the individual hydroxyl groups about the C–O bond.

The most readily available of the long-chain secondary alcohols are those having the hydroxyl group on the central carbon atom, and compounds of this type were therefore used in the initial experiments mentioned above. The properties of long-chain compounds are known to vary considerably, however, with the position of the polar group in the carbon chain (Oldham and Ubbelohde 1939) and it therefore seemed desirable to make dielectric measurements with various unsymmetrical secondary alcohols. For this programme, five different alcohols of molecular chain-length 20 carbon atoms were prepared (Churchward *et al.* 1950), having the hydroxyl groups in the 2-, 4-, 6-, 8-, and 10-positions

---

* Division of Industrial Chemistry, seconded to the Division of Electrotechnology, C.S.I.R.O., University Grounds, Sydney.

respectively. Normal chemical methods were used in the preparation and the product investigated therefore consisted, in each case, of the optically inactive mixture of the *d*- and *l*- isomers. The results of some dielectric measurements of these materials are given in this paper, and are compared with those previously obtained for the symmetrical compounds.

II. MATERIALS

The preparation of the *dl*- forms of the 2-, 4-, 6-, 8-, and 10-*n*-eicosanols has been previously described (Churchward *et al.* 1950). The melting points (uncorrected) are set out in Table 1.

TABLE 1

Compound	Melted Form (°C.)	Recrystallized from Ethanol (°C.)
dl-n-Eicosan-2-ol	58·4–58·8	58·5–59·0
dl-n-Eicosan-4-ol	53·8–54·2	55·0–55·3
dl-n-Eicosan-6-ol	55·2–55·4	57·7–57·9
dl-n-Eicosan-8-ol	54·9–55·2	57·7–57·9
dl-n-Eicosan-10-ol	64·8–65·6	65·3–65·6

III. APPARATUS AND PROCEDURE

The materials for measurement were pressed in the form of disks according to the method previously described (Meakins and Sack 1951). Because of the very high loss angles and dielectric constants of several of the *n*-eicosanols, however, it was found desirable to prepare thicker disks, using about 4·5 g. of material.

The apparatus used for the electrical measurements was the same as that for the previous programme. The temperature during measurement was 25±2 °C.

Except where otherwise stated, measurements were made within 2 hours of preparation of the samples.

IV. RESULTS AND DISCUSSION

(*a*) *Dielectric Properties of the Melted Forms*

(i) *Variation with the Position of the Hydroxyl Group in the Carbon Chain.*—In considering the dielectric properties of the *dl*-*n*-eicosanols, useful conclusions can be drawn from a comparison of the results for the 10-compound, which is the most nearly symmetrical, with those for the previously investigated symmetrical alcohol 10-nonadecanol (Meakins and Sack 1951). These two materials differ in molecular structure by only one methylene group, thus :

10-Nonadecanol : $CH_3(CH_2)_8CHOH(CH_2)_8CH_3$,

dl-n-Eicosan-10-ol : $CH_3(CH_2)_8CHOH(CH_2)_9CH_3$.

A further point of difference is the presence of the two optically active forms in *dl-n*-eicosan-10-ol, whereas 10-nonadecanol is a single compound. Neither the slight asymmetry of the *n*-eicosan-10-ol molecule nor its presence as two optically active isomers appears to have any pronounced effect on the dielectric properties, however, since there is no considerable difference in the loss factor and dielectric constant curves of the two compounds (see Fig. 1). The higher frequency range of the absorption in 10-eicosanol can be attributed to its larger molecular weight, this being in agreement with the results of the previous investigation.

The largest dielectric absorption in the *dl-n*-eicosanols is obtained with the 4-, 6-, and 8-compounds (see Fig. 2). In discussing this, it is of interest to refer to an investigation by Oldham and Ubbelohde (1939) of the thermal properties of a series of ketones having the polar group attached at different points in a chain of 17 carbon atoms. It was concluded from this work that molecular vibration is a maximum at points removed from the centre or ends of the carbon chain. This would apply particularly to the 4-, 6-, and 8-positions in the *dl-n*-eicosanols studied in the present program. Thus it appears that increased thermal agitation of the molecules as a whole favours the formation of hydrogen bonds on solidification (Meakins and Sack 1951), but also the bonds are formed more easily when the hydroxyl group is located at the part of the molecule having maximum vibration in the solid state.

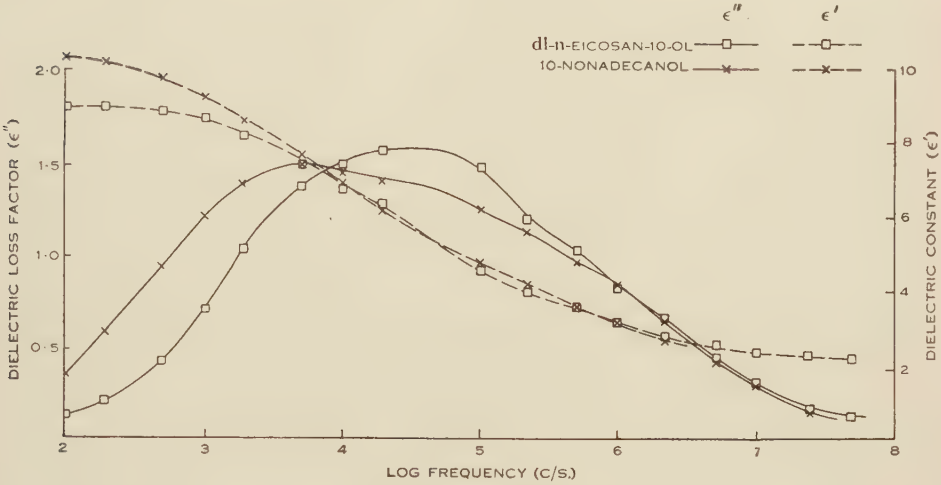


Fig. 1.—Comparison of the dielectric properties of *dl-n*-eicosan-10-ol ($C_{20}H_{42}O$) and 10-nonadecanol ($C_{19}H_{40}O$).

High dielectric absorption in the *dl-n*-eicosanols is accompanied by large increases in the dielectric constant towards lower frequencies (see Fig. 3). For 8-eicosanol ε′ reaches a value of 15 which is 50 per cent. higher than that previously observed for the symmetrical alcohol 10-nonadecanol. It has already been mentioned (Meakins and Sack 1951) that such values are abnormally large

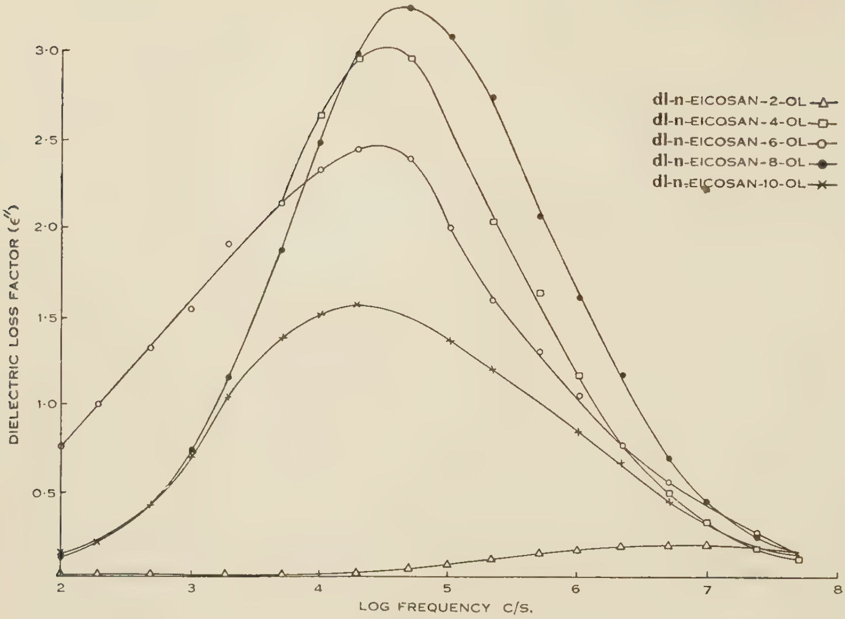


Fig. 2.—Dielectric loss factor of *dl-n*-eicosanols, melted form, measured within 2 hours.

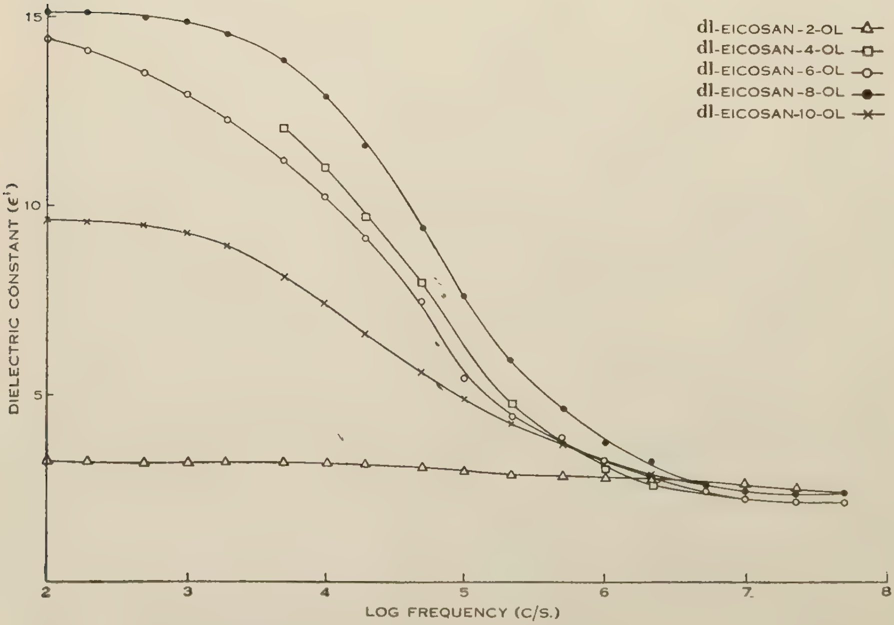


Fig. 3.—Dielectric constants of *dl-n*-eicosanols, melted form, measured within 2 hours.

compared with those for liquid alcohols, the effect probably being due to the development of the hydrogen-bond chains in straight lines, compared with a random development in liquids.

Thus far no mention has been made of the results obtained with *dl*-*n*-eicosan-2-ol. As shown in Figure 2, its dielectric absorption is very much smaller, and at higher frequencies, than for the other compounds of this series and the dispersion of the dielectric constant is correspondingly less (Fig. 3). This is not unexpected since long-chain compounds having a polar group in the 2-position have frequently been found to possess anomalous properties. X-ray measurements show that in such compounds the molecules in the crystal lattice are alternately reversed, end-to-end (Malkin 1931 ; Oldham and Ubbelohde 1939). Presumably, in the process of crystallization, polar groups at the outer ends of the molecules in an already established molecular layer attract the polar ends of molecules about to settle on the surface. It has already been suggested (Meakins and Sack 1951) that the formation of crystals in secondary alcohols involves competition between the intermolecular van der Waals forces and the forces due to hydrogen bonding. In *dl*-*n*-eicosan-2-ol, the latter would also be opposed by the polar forces of attraction between adjacent molecular layers and it is evident from the dielectric properties that this results in a molecular arrangement unsuitable for the development of extensive hydrogen-bond chains.

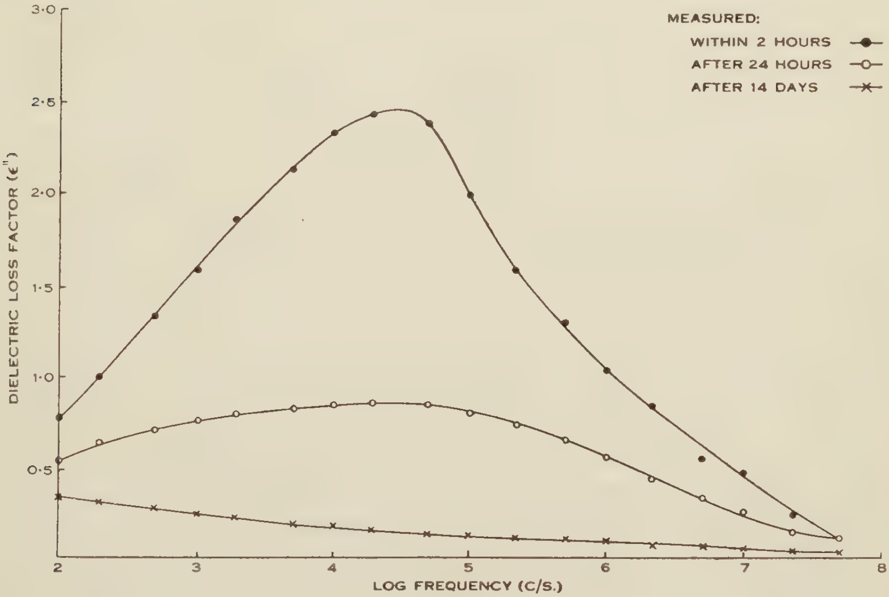


Fig. 4.—Change in dielectric absorption of *dl*-*n*-eicosan-6-ol on storing.

(ii) *Change in Dielectric Properties during Storage.*—In the previous investigation (Meakins and Sack 1951) it was found that symmetrical alcohols with less than 23 carbon atoms in the molecule showed considerable decreases in dielectric absorption during storage. Similar changes occur in the *dl*-*n*-eicosanols, presumably due to the same cause, namely, that the crystals formed at the melting point are unstable at room temperature and undergo a gradual molecular rearrangement leading to breaks in the hydrogen-bond chains.

With the 4-, 8-, and 10-compounds, the changes in dielectric absorption are of the same order as that previously observed for the symmetrical alcohol 10-nonadecanol. The 6-compound, however, gives much larger and more rapid changes, ε''_{max} falling to less than half its initial value within 24 hours and showing a further decrease after 14 days (see Fig. 4). With such a rapid rate of change, it is likely that an appreciable decrease occurs between solidification and measurement, and this may account for the dielectric absorption of *dl-n*-eicosan-6-ol being smaller (see Fig. 2) than would be expected from the favourable position of the hydroxyl group in the carbon chain (Section IV (*a*) (i)).

(*b*) *Dielectric Properties of the Recrystallized Forms*

The *dl-n*-eicosanols also behave similarly to the symmetrical alcohols in giving much smaller losses in the recrystallized than in the melted forms. The difference is similar to that observed for 10-nonadecanol, the ε''_{max} values for the recrystallized forms being only about 1 per cent. of those for the melted forms. These results indicate that the hydrogen-bond chains in the recrystallized forms are of comparatively small extent. As previously suggested (Meakins and Sack 1951) this is probably due to the low temperature of crystal formation which enables the van der Waals interchain cohesive forces to determine the molecular arrangement so that the spacing of the oxygen atoms is unsuitable for hydrogen bonding.

V. ACKNOWLEDGMENT

The authors wish to thank Mr. E. Filby, Division of Electrotechnology, C.S.I.R.O., for carrying out measurements at frequencies between 100 c/s. and 50 kc/s.

VI. REFERENCES

CHURCHWARD, V. R., GIBSON, N. A., MEAKINS, R. J., and MULLEY, J. W. (1950).—*J. Chem. Soc.* **1950** : 959.

MALKIN, T. (1931).—*J. Chem. Soc.* **1931** : 2796.

MEAKINS, R. J., and SACK, R. A. (1949).—*Nature* **164** : 798.

MEAKINS, R. J., and SACK, R. A. (1951).—*Aust. J. Sci. Res.* A **4** : 213.

OLDHAM, J. W. H., and UBBELOHDE, A. R. (1939).—*Trans. Faraday Soc.* **35** : 328.

THE DIELECTRIC PROPERTIES OF CRYSTALLINE LONG-CHAIN *n*-PRIMARY ALCOHOLS AT LOW FREQUENCIES

By R. J. Meakins* and Joan W. Mulley*

[*Manuscript received February* 23, 1951]

Summary

In a previous paper it was reported that certain crystalline forms of long-chain secondary alcohols show abnormally large dielectric absorption and dispersion of the dielectric constant at audio and radio frequencies. The present paper describes an extension of the investigation to *n*-primary alcohols and discusses the results in terms of the previously suggested theory of dielectric absorption due to the presence of hydrogen-bonded chains of hydroxyl groups.

The most significant feature of the results is that the absorption is larger and occurs at much lower frequencies than for the secondary alcohols. This is considered to be due to the end-to-end arrangement of the molecules in *n*-primary alcohol crystals, which gives double layers of hydroxyl groups and enables the formation of more extensive hydrogen-bond chains. The dielectric constants at the lowest frequencies are almost as large as those previously reported by Hoffmann and Smyth (1949) for the " waxy " forms of *n*-primary alcohols.

Further aspects described are the changes in dielectric properties during storage, the differences between the melted and recrystallized forms, and the effect of dispersing the molecules in an inert solid solvent. In each of these aspects the observed behaviour is similar to that previously reported for the secondary alcohols.

I. Introduction

It was recently reported (Meakins and Sack 1949, 1951; Meakins and Welsh 1951) that certain crystalline forms of long-chain secondary alcohols give abnormally large dielectric absorption and dispersion of the dielectric constant at audio and radio frequencies. This could be explained by the presence of hydrogen-bonded chains of hydroxyl groups capable of reversing their directions. A similar mechanism could also occur in *n*-primary alcohols and it therefore seemed of interest to extend the investigation to these compounds.

Some results of dielectric measurements of *n*-primary alcohols have already been reported (Baker and Smyth 1938; Hoffmann and Smyth 1949) but this work dealt chiefly with the " waxy " form stable at temperatures just below the melting point. The dielectric constants of primary alcohols in this form are larger than in the liquid, which the authors consider to be due to a Maxwell-Wagner type of polarization resulting from the presence of a non-conducting β_2-phase interspersed with a highly conducting α-phase. An alternative explanation is, however, that the high dielectric constant depends on the presence of hydrogen-bond chains which, on crystallization from the melt, change from a random to a unidirectional orientation. Such a mechanism was previously

---

* Division of Electrotechnology, C.S.I.R.O., University Grounds, Sydney.

proposed (Meakins and Sack 1951) to account for the high dielectric constants of crystalline long-chain secondary alcohols in relation to that of liquid alcohols.

The present paper gives the results of dielectric measurements of four *n*-primary alcohols at ordinary room temperatures. The work described deals chiefly with the " melted " form (β_2-phase) obtained by allowing the liquid alcohol to solidify and then cool down to about 20 °C. Some reference is also made, however, to the " recrystallized " form produced by crystallization from a solvent. The results are discussed with reference to the previously proposed mechanism of dielectric absorption due to the presence of reversible chains of hydroxyl groups.

II. MATERIALS

n-Hexadecyl alcohol ($C_{16}H_{33}OH$), m.p. 48·5 °C., and *n*-octadecyl alcohol ($C_{18}H_{37}OH$), m.p. 58·2 °C., were obtained from commercial cetyl alcohol by fractional distillation at about 2 mm. pressure through a 90 cm. column packed with monel-metal gauze saddles. The resulting products showed high electrical conductivity, probably due to the presence of oxidative degradation products (Hamon and Meakins 1950). The latter were removed by recrystallization from methanol and the crystals dried by heating at 35 °C. for several hours in a high vacuum.

n-Docosyl alcohol ($C_{22}H_{45}OH$), m.p. 70·5 °C., was prepared by the hydrogenation of ethyl behenate at 200 atmospheres pressure in the presence of a copper chromite catalyst, the product being purified by recrystallization from ethanol.

n-Hexacosyl alcohol ($C_{26}H_{53}OH$), m.p. 79·8 °C., was provided in the form of its acetate by Mr. K. E. Murray,* having been prepared by the vacuum fractionation of the acetates of commercial ceryl alcohol. The acetate was hydrolysed in this laboratory with 5 per cent. alcoholic potassium hydroxide solution and the product purified by recrystallization from ethanol.

III. APPARATUS AND PROCEDURE

The materials for measurement were prepared in the form of disks, 5 cm. in diameter, and approximately 0·2 cm. in thickness, by the pressing technique previously described (Meakins and Sack 1951).

The electrical measurements in the frequency range 50 c/s. to 50 kc/s. were made by a bridge method as previously used for the long-chain secondary alcohols. At the lower frequencies, i.e. between 0·4 and 4 c/s., a Schering bridge was used, the alternating potentials being supplied by a motor-driven sine-cosine potentiometer and 135 V. dry battery.

The samples were measured between mercury electrodes with a clamped steel guard ring. The measurements were made within about one hour of preparation of the sample except where otherwise stated. The temperature during measurements was 20 ± 0·5 °C.

---

* Division of Industrial Chemistry, C.S.I.R.O.

IV. Results and Discussion

(*a*) *Melted Forms of the* n-*Primary Alcohols*

(i) *Comparison with Secondary Alcohols.*—The dielectric absorption curves for the four *n*-primary alcohols are given in Figure 1. The most striking feature of the results is the low frequency of the absorption compared with that for the secondary alcohols (Meakins and Sack 1951) which usually give maxima at, or above, 10^4 c/s. This is probably related to the end-to-end arrangement of the molecules in *n*-primary alcohol crystals (Malkin 1931 ; Wilson and Ott 1934), which results in double layers of hydroxyl groups, compared with single layers in secondary alcohols. Such a structure would enable the formation of longer hydrogen-bond chains since, if a hydroxyl group in one layer was unsuitably placed for hydrogen bonding the chain could continue by bonding across to the next layer.

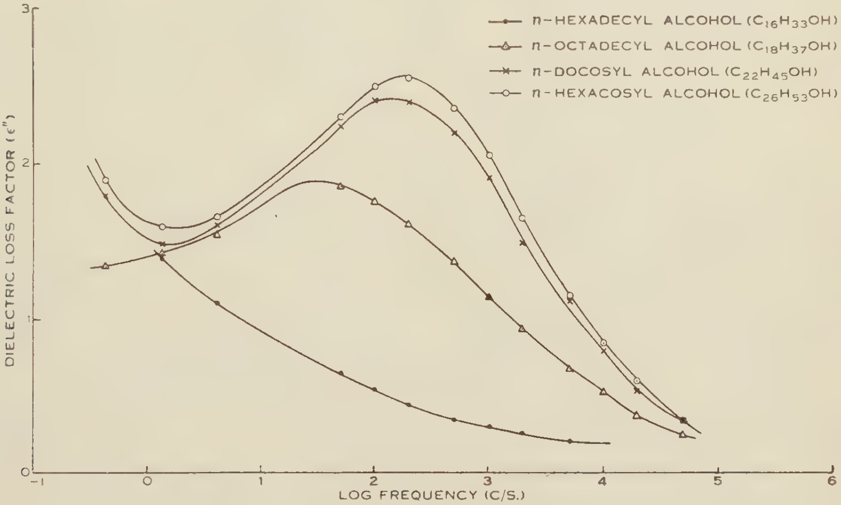


Fig. 1.—Dielectric absorption of *n*-primary alcohols—melted forms.

The presence of longer hydrogen-bond chains in the *n*-primary alcohols is also indicated, particularly with the C_{22} and C_{26} compounds, by the larger absorption (Fig. 1) and larger dispersion of the dielectric constant (see Fig. 2), compared with secondary alcohols of similar molecular weight. Thus, for *n*-hexacosyl alcohol ($C_{26}H_{53}OH$), ε' reaches a value of nearly 16 at the lowest frequency of measurement, and the shape of the curve suggests that it increases still further at lower frequencies.

(ii) *Comparison with Results of Hoffmann and Smyth.*—From these results it is evident that the high dielectric constants observed by Hoffmann and Smyth (1949) do not occur only in the waxy forms of the *n*-primary alcohols. Thus, with *n*-docosyl alcohol ($C_{22}H_{45}OH$) in the present program ε' reaches a value of 15 at 0·4 c/s., i.e. at the lowest frequency of measurement. This is nearly equal to the highest value (15·9) observed by Hoffmann and Smyth (1949) for the waxy form of this compound at a frequency of 500 c/s. The chief difference between the two forms which is apparent from these results is that the dielectric dispersion for the waxy form is displaced to frequencies more than one thousandfold higher than that for the melted form (β_2-phase).

Hoffmann and Smyth's result for the waxy form was obtained at a temperature of nearly 70 °C., where this type of structure is stable. Their measurements were continued, however, with gradual cooling to about 52 °C. where the conversion of the waxy form to the β_2-phase should be complete. It is interesting to note that the values given by these authors for the β_2-phase of n-docosyl alcohol at 52 °C. are in moderately good agreement with those obtained in the present program at 20 °C. With n-octadecyl alcohol the agreement is even better. Evidently, the abovementioned large difference in the frequencies of the dielectric dispersion for the waxy form and the β_2-phase is due to the different structures rather than to the difference in the temperatures of measurement.

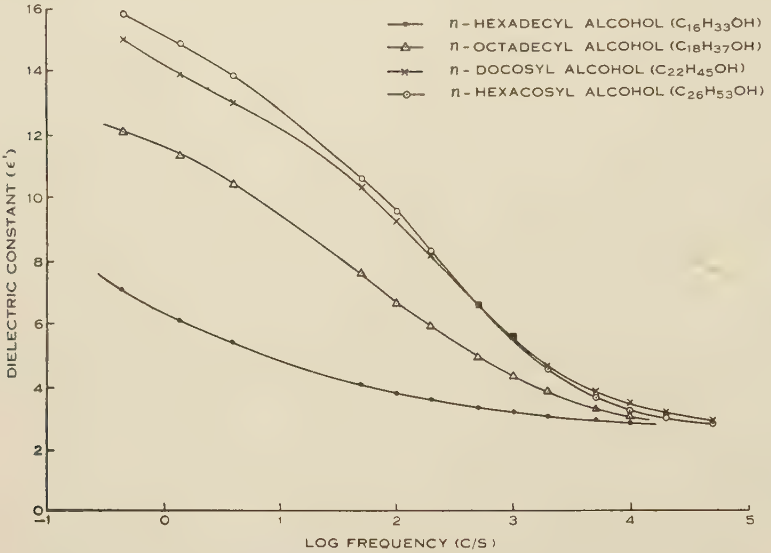


Fig. 2.—Anomalous dispersion of the dielectric constant for n-primary alcohols—melted forms.

(iii) *Variation of Dielectric Properties with Molecular Weight*.—With n-hexadecyl alcohol ($C_{16}H_{33}OH$), the lowest member of the present series, the dielectric properties for different sample disks were found to vary considerably, possibly due to its comparatively rapid rate of change after solidification (see Section IV (*a*) (iv)). The curves shown in Figures 1 and 2 are for a sample having loss factor and dielectric constant values intermediate between the two extremes in a series of six sample disks. The results were found to be similar at the higher frequencies but diverged considerably with decreasing frequency, the total variation in dielectric loss factor being about 2 : 1 at 50 c/s. A separate investigation is being made of the low frequency dielectric properties of *n*-hexadecyl and the lower *n*-primary alcohols.

In the results shown in Figure 1 for the higher homologues there is evidence of two main regions of absorption, largely superimposed in *n*-octadecyl alcohol ($C_{18}H_{37}OH$) but more clearly defined in *n*-docosyl ($C_{22}H_{45}OH$) and *n*-hexacosyl alcohols ($C_{22}H_{45}OH$). The reason for this is not apparent but it is evident that both absorption regions depend on the juxtaposition of hydroxyl groups since they are both eliminated by dispersal of the molecules in solid solution in a non-polar solvent (Section IV (*c*)).

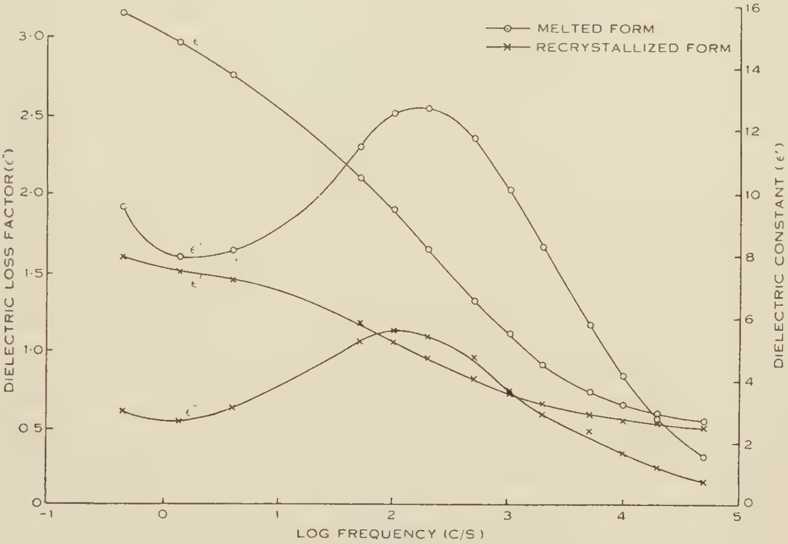


Fig. 3.—Dielectric properties of the melted and recrystallized forms of *n*-hexacosyl alcohol ($C_{26}H_{53}OH$).

An important feature of the results given in Figure 1 is the large displacement of the dielectric absorption towards higher frequencies with increasing molecular weight. This is most marked in the range of molecular chain-length between 16 and 22 carbon atoms, with only a small further displacement for the C_{26} compound. A similar region of rapid displacement was observed in the secondary alcohols (Meakins and Sack 1951) between 19 and 23 carbon atoms. As previously suggested the higher frequency range of the absorption for the compounds of higher molecular weight could be due to the relatively smaller effect of the hydroxyl groups in establishing a molecular arrangement in the crystal suitable for the development of long hydrogen-bond chains.

(iv) *Change in Dielectric Properties during Storage at Room Temperature.*—The rapid rate of change in the dielectric properties of *n*-hexadecyl alcohol during storage at room temperature has already been mentioned as a probable cause of variation in the results for different samples (Section IV (*a*) (iii)). With the higher members of the series the changes occur much more slowly. Thus, at 50 c/s., the loss factor of a *n*-hexadecyl alcohol sample decreased to about half its initial value within 24 hours whereas the corresponding decrease for *n*-octadecyl alcohol was only 10 per cent. After 14 days the former decreased by a further 5 per cent. and the latter by about 4 per cent. The two higher homologues showed even smaller changes than *n*-octadecyl alcohol. Presumably, with the larger molecules, molecular diffusion is more difficult, and rearrangements leading to the breaking of hydrogen-bond chains occur much more slowly.

(*b*) *Recrystallized Forms of the* n-*Primary Alcohols*

A comparison of the dielectric loss factor and dielectric constant curves for the melted and recrystallized forms of *n*-hexacosyl alcohol is given in Figure 3.

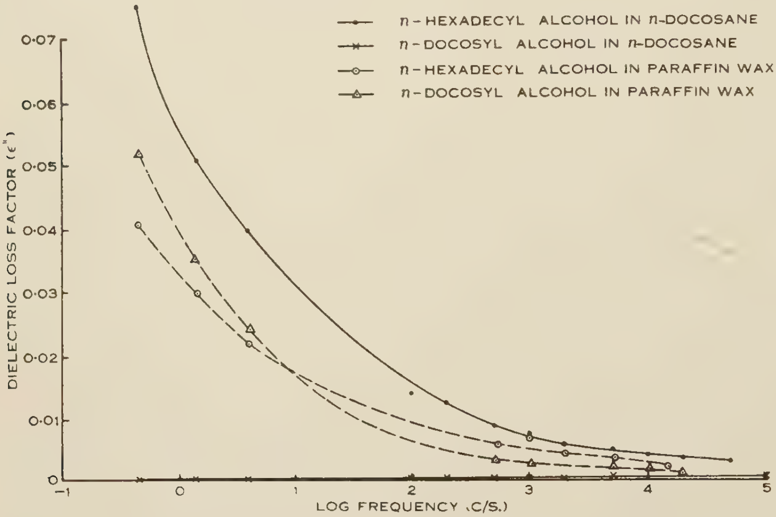


Fig. 4.—Dielectric absorption of 0·1m systems of *n*-primary alcohols in *n*-docosane and paraffin wax.

A similar relationship between the two forms was observed with *n*-docosyl and *n*-octadecyl alcohols. As mentioned above, the results for *n*-hexadecyl alcohol were variable, but the difference between the two forms was certainly not as large as with the secondary alcohols of lower molecular weight. The larger absorption of the recrystallized forms of the *n*-primary alcohols in relation to the secondary alcohols may be due to the greater freedom of orientation of the hydroxyl groups at the ends of the paraffin chains. This would facilitate hydrogen bonding even in the presence of strong van der Waals forces.

(*c*) *Dilute Solid Systems of* n-*Primary Alcohols with* n-*Docosane and with Paraffin Wax*

Previous dielectric research with solid mixtures of organic long-chain compounds has shown that the properties of such mixtures are largely determined by the relative number of carbon atoms in the molecules of the solvent and solute. In the present programme, the hydrocarbon docosane ($C_{22}H_{46}$) was used as a non-polar solvent and the systems chosen so that, in one case (*n*-hexadecyl alcohol, $C_{16}H_{33}OH$) the molecular chain-length of the solute was considerably shorter, and in the other case (*n*-docosyl alcohol, $C_{22}H_{45}OH$) it was equal to that of the solvent. The results (see Fig. 4) show that the former gives high dielectric absorption such as would be expected if the cetyl alcohol formed a separate phase. With the latter the absorption is negligibly small over the whole frequency range of the measurements, suggesting that the *n*-docosyl alcohol molecules are dispersed individually in the crystal lattice of the *n*-docosane (i.e. forming a solid solution), thus breaking the hydrogen-bond chains.

These results suggest that in systems of *n*-primary alcohols with paraffins, solid solutions are formed only when the molecular chain-lengths of the solute and solvent are similar. This is in agreement with the results of previous investigations of systems of other types of long-chain polar compounds with paraffins (Meakins 1949 ; Meakins and Sack 1949, 1951). In each case, when the chain-lengths are considerably different the dielectric properties are similar to the pure polar compound and suggest the formation of separate phases.

Figure 4 also shows the results of dielectric measurements of systems of *n*-hexadecyl and *n*-docosyl alcohols with paraffin wax. Both systems give considerable dielectric absorption, which is to be expected since the average molecular chain-length of the hydrocarbons in the wax is about 29 carbon atoms, and the two ingredients of each system would therefore exist largely as separate phases. The discrepancy between the dielectric absorption results for *n*-hexadecyl alcohol in its two systems may be due to the paraffin wax being a mixture of hydrocarbons of different chain-lengths which would result in different conditions of crystallization.

V. ACKNOWLEDGMENTS

The authors wish to thank Mr. B. V. Hamon and Miss M. Clark, Division of Electrotechnology, C.S.I.R.O., for making the electrical measurements.

VI. REFERENCES

BAKER, W. O., and SMYTH, C. P. (1938).—*J. Amer. Chem. Soc.* **60** : 1229.

HAMON, B. V., and MEAKINS, R. J. (1950).—*Nature* **166** : 29.

HOFFMANN, J. D., and SMYTH, C. P. (1949).—*J. Amer. Chem. Soc.* **71** : 431.

MALKIN, T. (1931).—*J. Chem. Soc.* **1931** : 2796.

MEAKINS, R. J. (1949).—*Aust. J. Sci. Res.* A **2** : 405.

MEAKINS, R. J., and SACK, R. A. (1949).—*Nature* **164** : 798.

MEAKINS, R. J., and SACK, R. A. (1951).—*Aust. J. Sci. Res.* A **4** : 213.

MEAKINS, R. J., and WELSH, H. (1951).—*Aust. J. Sci. Res.* A **4** : 359.

WILSON, D. A., and OTT, H. (1934).—*J. Chem. Phys.* **2** : 231.

SOME ELECTROPHORETIC STUDIES OF PROTEINS IN THE ADSORBED STATE

By C. W. N. Cumper* and A. E. Alexander*

[*Manuscript received February 2, 1951*]

Summary

The electrophoretic mobility of bovine γ-globulin, haemoglobin, and insulin adsorbed upon particles of different chemical types (mostly hydrocarbon derivatives) has been studied as a function of protein concentration and of pH.

On the majority of the surfaces used the isoelectric points of the bovine γ-globulin and insulin were 6·0 and 5·0 respectively. However, with certain particles, especially cetyl alcohol, a marked shift in the isoelectric point was observed and on either side of the isoelectric point the mobility depended to a certain extent upon the nature of the particle. These effects have been ascribed to two causes: (*a*) a specific reaction between the protein and the particle and (*b*) an incomplete coverage of the particle by the protein.

Insulin adsorbed upon particles became more negatively charged on standing, probably due to the insulin dissociating into smaller units.

I. Introduction

Electrokinetic phenomena, in particular the measurement of electrophoretic mobility, have played an outstanding part in the study of colloids, both of the hydrophilic and hydrophobic types. The development of the Tiselius moving boundary method, applied with such success in the analysis and separation of protein mixtures, has led to some tendency to belittle the value of the older method of studying hydrophilic colloids by adsorption on inert particles. A more reasonable perspective is provided by two recent articles by Booth (1948) and Overbeek (1950) who stress the importance of precise mobility measurements upon free and adsorbed proteins in order to test certain aspects of the general theory of electrophoresis. In addition, electrophoretic studies of adsorbed protein films offer one means of approach to the problem of "surface denaturation"—a problem of considerable importance in connection with protein-stabilized foams and emulsions and of potential biological interest in connection with immunological reactions.

Many native proteins, such as the ones considered here, possess a corpuscular shape and a considerable solubility in water or dilute salt solutions. These properties suggest that the constituent polypeptide chains are folded up in some way which masks an appreciable proportion of the side-chains, the system being thermodynamically unstable with respect to the more extended forms which arise on "denaturation". Changes in shape brought about by surface denaturation might thus alter the structure and amino-acid composition of

---

* Chemistry Department, N.S.W. University of Technology, Broadway, Sydney.

the surface of the protein molecule, leading to changes in the electrophoretic mobility, a property susceptible to accurate measurement.

Direct comparison of the electrophoretic mobility of proteins in the dissolved and in the adsorbed states was first carried out by Abramson (1932) (see also Abramson, Moyer, and Gorin 1942). He compared serum albumin adsorbed upon particles of quartz and collodion with Tiselius's data for solutions of this protein, finding complete agreement over a pH range of 1½ units on either side of the isoelectric point. Similar agreement was found with pseudoglobulin, but with egg albumin the isoelectric point was somewhat higher in the adsorbed state (pH 4·82 as compared with 4·55), although the mobility-pH curves were essentially parallel. Completely surface denatured egg albumin gave a shift of the isoelectric point in the same direction, but greater in magnitude (to pH 5·02).

Moyer (1938), also working with egg albumin, found that ionogenic surfaces such as aluminium oxide or silica gel had a noticeable effect, not upon the isoelectric point, but upon the mobilities at other pH values. He ascribed this to a change in the effective radius of the particles, presumably due to hydration. It cannot be due to a chemical interaction since we would then expect a shift in the isoelectric point.

Many other examples of the comparison of the isoelectric points determined by the two electrophoretic techniques may be found in the literature. Some show agreement ; others, such as chymotrypsin and chymotrypsinogen (Anderson and Alberty 1948), a wide divergence.

In most electrophoretic studies of adsorbed proteins the particles used have been solids of complex chemical nature, such as collodion, carbon, glass, etc. The precise nature and the density of the polar groups at the surface is thus uncertain, and being solids the surface energies cannot readily be measured. To overcome these limitations we have employed particles made from a variety of paraffin derivatives (alcohols, acids, amides, hydrocarbons, etc.) in both the liquid and solid states. In these systems the nature and density of the polar groups at the surface are known, and with the liquid compounds the surface energy is accessible to direct measurement.

II. EXPERIMENTAL

The suspensions used in the electrophoresis experiments were all prepared by the method of Powney and Woods (1940). A suspension containing about 0·1 per cent. of the organic compound, dispersed as particles of about 1–3 μ diameter, is obtained.

Two proteins were chiefly studied in this electrophoretic work, namely crystalline bovine γ-globulin (from Armour and Co.) and crystalline insulin (one sample from Boots Ltd., and another from Armour and Co.). Their solubility relations are similar, being practically insoluble in pure water, but soluble in dilute salt solutions and precipitated by high concentrations of salt. Bovine γ-globulin has a molecular weight of 150,000, and can be denatured and coagulated quite readily. Insulin has a molecular weight of 48,000 in concentrated solutions at pH 7 but, unlike the γ-globulin, it is one of the few proteins that cannot readily be denatured. Insulin has the further property of dissociating reversibly, in dilute solution, into four units of molecular weight 12,000 (Gutfreund 1948). A few experiments were also performed upon horse haemoglobin, for which we are indebted to Dr. Gutfreund. Acetate buffers were used for the pH range 3·8 to 5·7, and phosphate for the range pH 5·8 to 8·0.

The electrophoretic mobilities were measured in an all-glass two-tube type of capillary cell, immersed in a water thermostat and fitted with silver-silver chloride electrodes (Alexander and Saggers 1948). The reproducibility of the results obtained with this apparatus is extremely good—about ±1 per cent. in favourable circumstances.

Interfacial tensions were obtained by the ring method, using the simple technique described earlier (Alexander 1947). The apparatus was thermostatted at 25 °C., the same temperature as used in the electrophoretic studies.

III. RESULTS

(a) *The Adsorption of Bovine γ-Globulin*

(i) *The Effect of Protein Concentration.*—In the following experiments the suspension was added to the buffered solution of the protein: Abramson in most of his work mixed the suspension and protein first, diluted the system, and finally brought it to the required ionic strength. To 8 ml. of an M/50 phosphate –NaOH buffer (pH 7·0), 1 ml. of the protein solution and 1 ml. of the suspension were added. The mobilities at protein concentrations between 10^{-2} and 10^{-6} per cent. were found with various particles, whose mobilities in the absence of the protein lay between −0·53 and −2·93 μ/sec./volt/cm. As shown in Figure 1 all the suspensions, with the sole exception of cetyl alcohol, gave mobilities between −0·41 and −0·45 μ/sec./volt/cm. at protein concentrations greater than 0·001 per cent.

It seems, therefore, as though the surface of the particle is just saturated at a protein concentration of 0·001 per cent. and, assuming that the particles are 2μ in diameter and that all the protein is adsorbed, this corresponds to about 1 mg. of protein per sq. metre of surface. This is close to the area occupied by spread monolayers of proteins when close-packed at an interface. Since the mobility does not change at higher concentrations then, either polymolecular films are not being built up, or the layers subsequently adsorbed present an essentially identical surface towards the aqueous phase.

Bovine γ-globulin adsorbed upon cetyl alcohol particles gave a mobility of −0·83 μ/sec./volt/cm. at the concentration of 0·001 per cent., and the same value at 0·10 per cent. The protein adsorbed upon particles made from white oil saturated with cetyl alcohol and from oleyl alcohol behaved normally. This may be due to these surfaces being liquid whereas cetyl alcohol particles are solid (see page 377).

The effect of protein concentration upon the interfacial tension is also shown in Figure 1. The tension was measured at intervals until it became reasonably constant: at the higher protein concentrations this usually took about 2 hours. When skins were present the surface showed marked elastic properties, thus introducing an unknown error into the measurements, although they are still characteristic of the system under test.

The lowering of the interfacial tension, as would be expected, is much greater for the interfaces of high initial energy.

(ii) *Mobility-pH Curves.*—The mobility-pH curves over the pH range 3·6 to 8·0 were found for bovine γ-globulin adsorbed upon a variety of surfaces. The measurements were carried out at 25 °C., with a protein concentration of 0·01 per cent. and an ionic strength of 0·04.

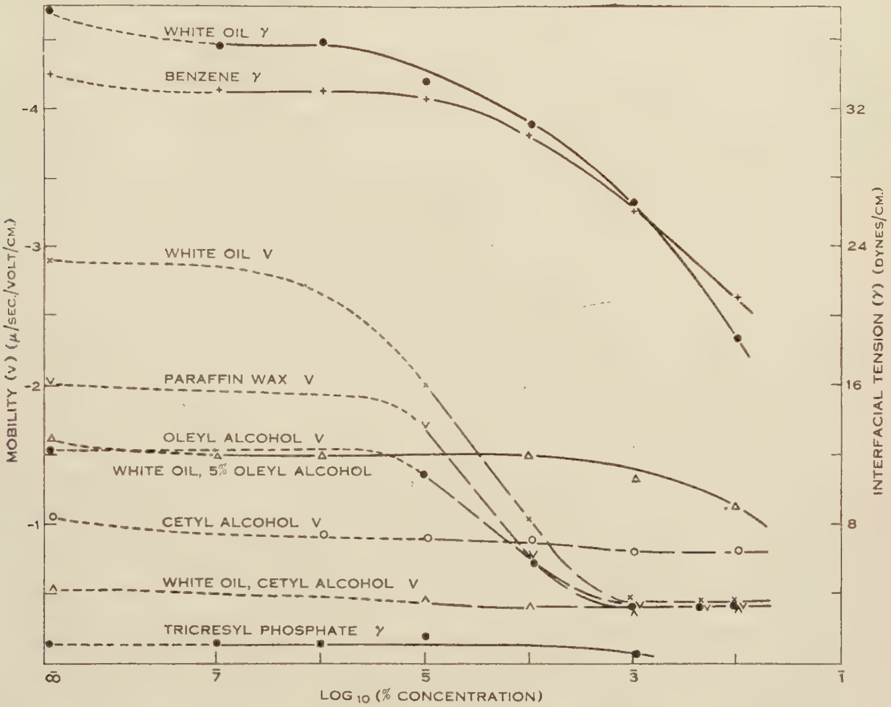


Fig. 1.—The adsorption of bovine γ-globulin upon a variety of interfaces at pH 7·0. Electrophoretic mobility and interfacial tension as a function of protein concentration.

As seen from the results shown in Figure 2 and Table 1 several curves show an isoelectric point of pH 5·97 ±0·02, in good agreement with the accepted value of 6·0. With a few surfaces, however, a shift in the isoelectric point was obtained, particularly with cetyl alcohol, which gave a value of *c.* 4·6 with the usual laboratory material, and a value of 5·1 with a highly purified specimen. Smaller shifts were found with particles of oleyl alcohol (pH 5·5), and white oil (pH 5·4). If, however, the protein concentration was increased from 0·01 per cent. to 0·05 per cent. the isoelectric points now found were 6·0 and 5·9 with particles of white oil and oleyl alcohol respectively, whilst the isoelectric point with cetyl alcohol particles was unchanged. At other pH values a small spread of mobilities with the different particles was found ; the further the pH from the isoelectric point the greater being the spread. This effect was more marked at acid than at alkaline pH's.

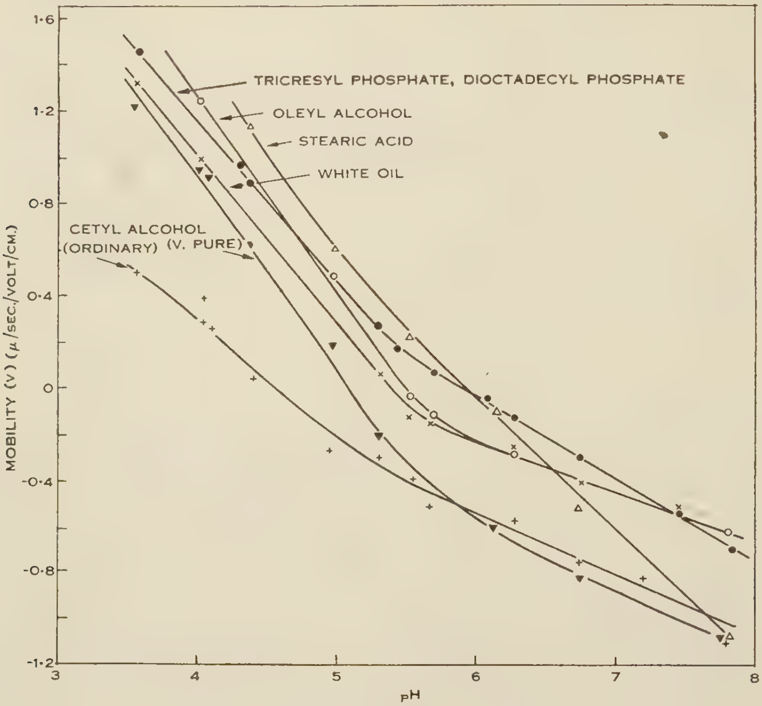


Fig. 2.—The adsorption of bovine γ-globulin upon a variety of interfaces. Electrophoretic mobility as a function of pH.

TABLE 1

Particle	pH				
	4·0	5·0	6·0	7·0	8·0
Cetyl alcohol (solid)	0·33	—0·19	—0·52	—0·79	—1·07
Very pure cetyl alcohol (solid) ..	0·93	0·07	—0·54	—0·85	—1·12
Ethyl stearate (liquid)	1·17	0·57	—0·02	—0·40	—0·69
Oleyl alcohol (liquid)	1·27	0·42	—0·21	—0·45	—0·63
Paraffin wax (solid)	1·53	0·82	0·00	—0·40	—0·67
Stearamide (solid)	1·15	0·52	—0·02	—0·42	—0·72
Stearic acid (solid)	1·47	0·60	—0·03	—0·62	—1·13
Tricresyl phosphate (liquid) ..	1·28	0·54	—0·03	—0·43	—0·69
Tricresyl phosphate+dioctadecyl phosphate (liquid)	1·15	0·48	—0·04	—0·39	—0·72
White oil (liquid)	1·00	0·30	—0·22	—0·45	—0·60
White oil+dioctadecyl phosphate (liquid)	1·12	0·48	—0·03	—0·43	—0·64

Discussion

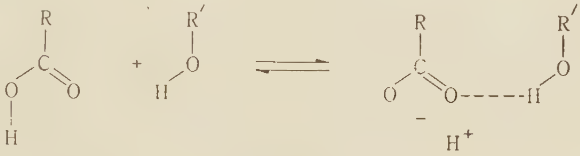
The spread of mobilities obtained with the different surfaces, similar to Moyer's results on egg albumin, provides further proof that this is not due to a specific reaction between the protein and the particle. It does not, however, confirm Moyer's theory that the effect is due to a change in the effective radius of the particle.

From the present work, considered in conjunction with our earlier studies on adsorbed films of proteins (Cumper and Alexander 1950), we believe that these apparent shifts in isoelectric point may arise in one (or both) of two ways :

(*a*) Interaction between the adsorbed protein film and polar groups on the particle's surface.

(*b*) Incomplete coverage of the surface of the particle by the protein film.

With cetyl alcohol particles, which appeared to lower the isoelectric point of bovine γ-globulin, the most likely interaction would be hydrogen bonding between the –OH and –COO⁻ groups of the protein. This, by increasing the stability of the ion relative to the undissociated acid,



would tend to increase the carboxylic dissociation constant and thus to move the isoelectric point to lower pH values. Some preliminary experiments using mixed particles of cetyl alcohol and stearic acid, and of cetyl alcohol and hexadecylamine, have shown that the alcohol can modify the mobility very considerably, even when the acid or amine is fully ionized. With stearic acid the shift is in the direction anticipated, but its magnitude (*c.* 0·25 pH unit) is less than found with proteins. It is planned to carry out a more thorough investigation of this aspect of the problem using monolayer techniques as well as electrophoresis.

As regards mechanism (*b*)—incomplete coverage of the surface—earlier work (Cumper and Alexander 1950) has shown that the thickness of adsorbed protein films depends greatly on the nature of the interface as well as upon the pH and ionic strength of the medium. In some systems the loss of protein by surface coagulation, and in others the small tendency to adsorb protein, would result in the interface being incompletely covered with a protein film and thus behaving electrophoretically as a composite surface. The particles considered here are all negatively charged at neutral pH, with isoelectric points in the region of 2–3 (Growney 1941) and would thus tend to lower the isoelectric point below that for the completely covered protein surface. With liquid surfaces of high energy, such as white oil, loss of protein by coagulation is probably the determining factor, for on increasing the concentration of bovine γ-globulin from 0·001 per cent. to 0·05 per cent. the normal isoelectric point was obtained. On the other hand with solid surfaces, particularly those of a hydrophilic nature, all the evidence points to the adsorbed protein film being very thin, particularly as we move away from the isoelectric point. A variation of film thickness with pH would provide a reasonable explanation for the rotation of the pH-mobility curves found in some cases.

(*b*) *The Adsorption of Horse Haemoglobin*

This protein was only studied at three surfaces. Paraffin wax and oleyl alcohol particles gave approximately the same isoelectric point of pH 6·1, while that adsorbed upon both the ordinary and the very pure cetyl alcohol gave a marked shift of about 1½ units to lower pH values. The experimental points obtained were, however, rather scattered. These results can probably be interpreted in the same manner as those for bovine γ-globulin.

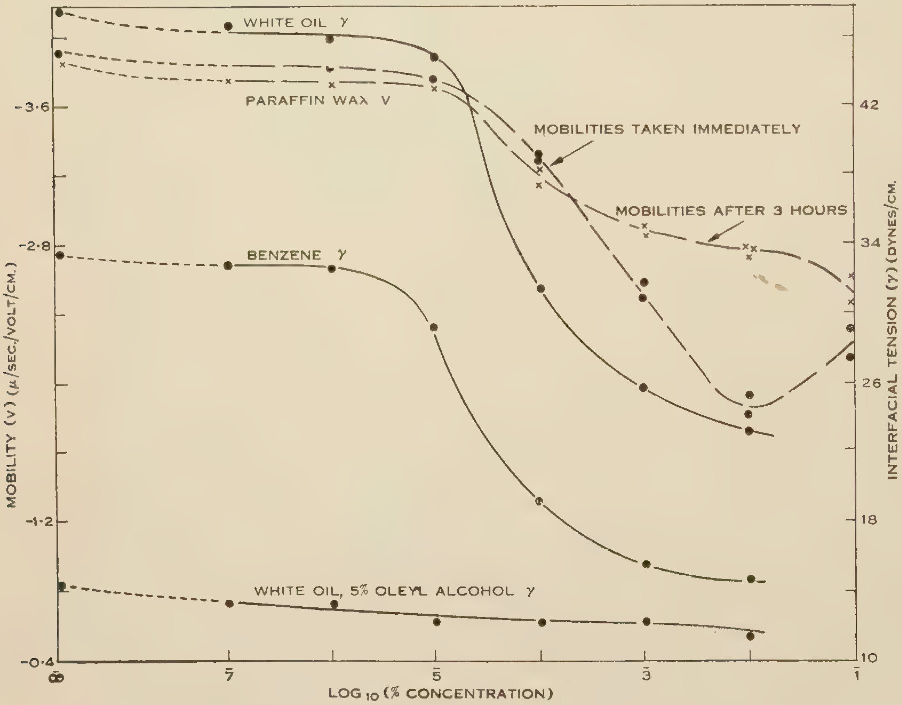


Fig. 3.—The adsorption of insulin at pH 7·0. Electrophoretic mobility and interfacial tension as a function of concentration.

(*c*) *The Adsorption of Insulin*

(i) *The Effect of Protein Concentration.*—Figure 3 shows the interfacial tension-concentration relations for a number of interfaces, and also the mobility for a paraffin wax suspension. As with bovine γ-globulin appreciable adsorption begins at a concentration of about 10^{-5} per cent., but unlike the globulin, the mobilities showed pronounced time effects. The great difference between mobilities taken immediately and after keeping for 3 hours can be seen from Figure 3. This phenomenon is discussed in Section III (*c*) (iii) below.

Although it is generally held that insulin cannot be denatured, definite indications of a skin were found at the interfaces of high energy, namely white oil and benzene. When the ring was pulled through the interface a film remained attached to it and, on breaking, it contracted rapidly to a visible skin which redissolved quite readily.

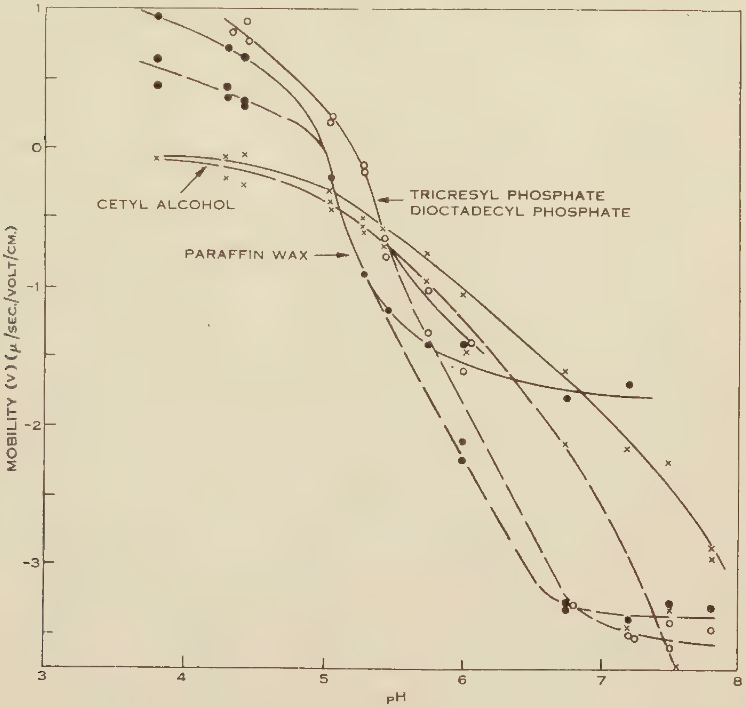


Fig. 4.—The adsorption of insulin upon a variety of interfaces. Electrophoretic mobility as a function of pH.

(ii) *Mobility-pH Curves.*—The mobility-pH curves showed the same effects of " rotation " and shift in the isoelectric point as did the bovine γ-globulin. In addition, however, marked time effects were found and, provided a standard procedure was followed, two quite distinct curves could usually be obtained by taking the mobilities immediately after mixing and after standing for say 3 hours. Except in the region of the isoelectric point the mobility increased on standing on the alkaline side and decreased on the acid side, showing that the particle was becoming more negatively charged. With some adsorbing surfaces (e.g. quartz) two distinct curves could not be obtained over the whole pH range, while with others (e.g. stearic acid) the change of mobility with time was small.

In the majority of cases an isoelectric point of pH $4 \cdot 98 \pm 0 \cdot 05$ was obtained (Fig. 4 and Table 2), but with insulin adsorbed upon particles made from a mixture of tricresyl phosphate and dioctadecyl phosphate an isoelectric point of

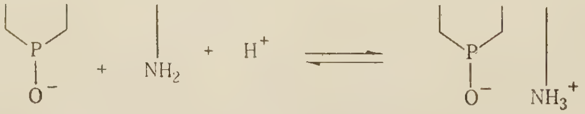
5·2 was obtained, while with cetyl alcohol the particles were always negatively charged, even at pH 3·8. The latter is probably mainly due to incomplete coverage of the surface, more marked in this case than with bovine γ-globulin owing to the less ready denaturation of insulin, as pointed out previously. There may also be a specific reaction with the insulin as discussed on page 377.

TABLE 2

Particles	pH											
	3·5		4·5		5·0		5·5		6·5		7·5	
	a	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*
Cetyl alcohol (pure) (solid)	−0·02	−0·02	−0·12	−0·17	−0·32	−0·37	−0·60	−0·75	−1·40	−1·82	−2·38	−3·5
Oleyl alcohol (liquid) ..	0·90	0·90	0·50	0·50	−0·02	−0·02	−0·95	−1·12	−1·47	−2·67	−1·50	−3·28
Paraffin wax (solid) ..	1·07	0·57	0·58	0·30	0·02	0·00	−1·22	−1·35	−1·75	−2·97	−1·75	−3·35
Quartz (solid)	1·10	1·10	0·65	0·64	0·02	0·00	−0·82	−0·85	−1·99	−2·07	−2·13	−2·7
Stearic acid (solid) ..	1·02	1·02	0·70	0·62	0·00	−0·02	−1·12	−1·27	−2·25	−2·50	−2·62	−2·80
Tricresyl phosphate + dioctadecyl phosphate (liquid)	1·08	1·08	0·75	0·75	0·28	0·23	−0·87	−0·90		−2·70		−3·50

a. Mobilities taken immediately.
b. Mobilities taken after 3 hours.

The shift of the isoelectric point to a higher pH when insulin was adsorbed upon particles of tricresyl phosphate containing dioctadecyl phosphate might conceivably be ascribed to an increase in the basic dissociation constants of the free amino groups consequent upon their juxtaposition to an ionized phosphate group, e.g.



Wintersteiner and Abramson (1933) in their work on the mobility of adsorbed insulin found an isoelectric point of 5·30–5·35 for precipitated, amorphous insulin and for insulin adsorbed onto quartz particles, and a value of 5·0 for insulin crystals. Their mobility-pH curves corresponded to the present results for the mobility taken immediately after mixing. Hall (1941*a*, 1941*b*) claimed that his results on dissolved insulin were in general agreement with those of Wintersteiner and Abramson, though he could not work near the isoelectric point, and his mobilities depended to a large extent on the previous treatment of insulin. Various samples of insulin are known to differ slightly in their physical properties and it seems as though the present sample (Boots Ltd.) had an isoelectric point 0·3 units lower. This lower value was also found in later work using a sample of insulin from Armour and Co., which also gave similar time effects.

In the above work of Wintersteiner and Abramson the insulin and quartz particles were first mixed and, after standing for 10 minutes, the buffer added. The mobilities, using this method and the present one of adding the particles last, were found to be identical.

(iii) *Investigation of the Time Effects.*—The increase in the negative charge on standing was observed, to varying extents, with all the particles tried. This cannot be due to a slow adsorption of the insulin as two quite distinct and continuous curves are obtained which only coincide at the isoelectric point.

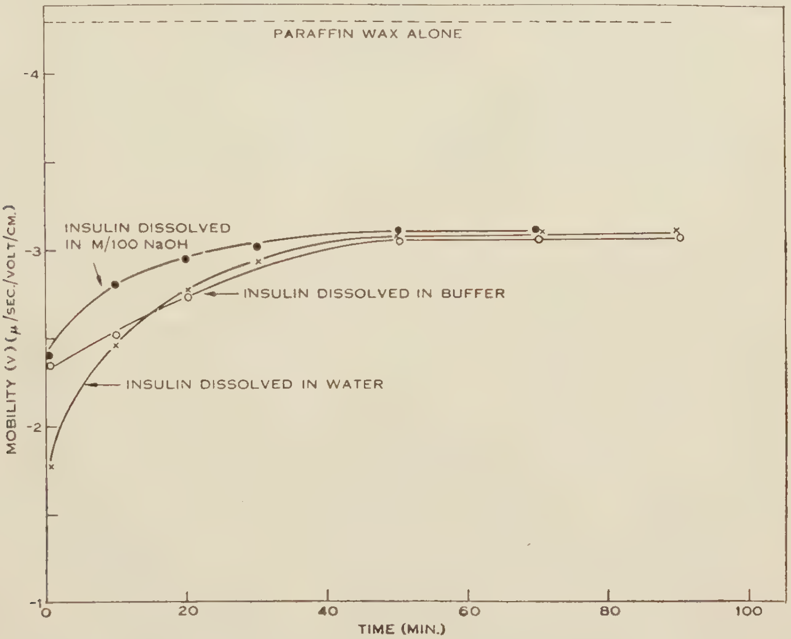


Fig. 5.—The adsorption of insulin upon paraffin wax at pH 6·75. Electrophoretic mobility as a function of time.

Also the mobility found after standing lies between that of the particles alone in the absence of insulin and the mobility taken immediately after mixing.

The variation of mobility with time was followed for insulin on paraffin wax particles at pH 6·75. As seen from Figure 5 the initial mobility varied with the medium in which the insulin was dissolved, although the final mobility was the same in each case.

The mobilities of insulin, at various concentrations, adsorbed onto a paraffin wax suspension at pH 7·0 were found immediately after mixing and after standing for 3 hours. The curves obtained (see Fig. 3) were anomalous but reproducible, except for the mobility taken immediately at a concentration of 0·1 per cent. The mobilities found after standing fell rapidly between insulin concentrations of 0·00001 and 0·001 per cent. In the concentration range of 0·005 to 0·05 per cent. the mobility was approximately constant but instead of remaining at this value as one would expect (cf. p. 374) it decreased again at a concentration of 0·1 per cent. The concentration at which the mobility of adsorbed γ-globulin became constant was 0·001 per cent. (cf. Fig. 1).

Up to a concentration of 0·0001 per cent. the mobility found immediately after mixing was practically the same as that after standing but it continued to fall rapidly up to 0·01 per cent., above which it increased again. The insulin concentration used in the mobility-pH work (0·01 per cent.) corresponds to this minimum mobility.

The effect of temperature upon these time effects was examined by taking two 0·1 per cent. solutions of insulin in M/50 phosphate buffer of pH 7·0 and keeping them overnight at 0 °C. and at 45 °C. respectively. To 1 ml. of these solutions 1 ml. of a paraffin wax suspension and 8 ml. of a salt solution were added so as to bring the ionic strength up to 0·04. The mobilities (Table 3) were taken immediately after mixing and, in addition, mixtures of the insulin and suspension were allowed to stand at these temperatures for 3 hours before the salt solution was added, and the mobilities then measured. (All mobilities were taken at 25 °C.)

Table 3

	Temperature of Adsorbing Insulin	
	0 °C.	45 °C.
Mobilities taken immediately	2·33 μ/sec./volt/cm. ↓ after 30 min. at 25 °C. 2·68	2·74 μ/sec./volt/cm. ↓ after 30 min. at 25 °C. 2·85
Mobilities after standing for 3 hours	2·74 ↓ after 30 min. at 25 °C. 2·95	2·95 ↓ after 30 min. at 25 °C. 2·95

The possibility that the time effects were due to traces of a surface active impurity in the insulin was examined experimentally. To a 0·1 per cent. solution of insulin at pH 8·0, quartz particles, sufficient to adsorb about 10 per cent. of the insulin, were added and left to stand overnight. The insulin remaining in solution still showed the same time effects as before, confirming that they are not due to a surface active impurity slowly displacing the insulin from the surface.

In the Introduction it was mentioned that the insulin molecule dissociates reversibly, in dilute solutions, into smaller units and it would appear that these time effects are connected with this phenomenon. The following factors seem to be involved :

(*a*) The degree of dissociation decreases with increasing concentration.

(*b*) The rate of dissociation of the insulin, both in solution and on the surface, may be slow, as is often the case with proteins. The molecular weight of insulin determined with a film balance would indicate dissociation on a surface into several units (Allan, personal communication).

(*c*) The surface activity of the dissociation fragments might well be greater than that of the parent molecule, since it is conceivable that dissociation would produce particles of greater asymmetry.

In view of the lack of experimental evidence on these points, however, we do not feel justified in analysing the matter in more detail. The results in Table 3 would indicate that an increase of temperature increases the dissociation and that further dissociation occurs in the presence of a surface.

IV. Conclusions

The present work clearly indicates that the mobility behaviour of adsorbed proteins, as studied by the microelectrophoretic technique, will only agree with that found by the Tiselius method under certain conditions. In the first place the adsorbing surface should not contain polar groups capable of interacting with specific groups in the protein ; secondly, the protein concentration must be adequate to ensure that the mobility is independent of it over the whole pH range. It would seem that a liquid paraffin most nearly meets the former requirements, although in cases of doubt it would be advisable to try other materials as well. Some of the apparent anomalies reported in the literature may have been due to the use of unsuitable particles.

V. Acknowledgments

We should like to thank Armour and Co. and Boots Ltd. for gifts of the proteins used in these investigations. One of us (C.W.N.C.) would also like to thank the Department of Scientific and Industrial Research for financial assistance and the Goldsmiths' Company for the award of a travelling scholarship.

VI. References

Abramson, H. A. (1932).—*J. Gen. Physiol.* **15** : 575.

Abramson, H. A., Moyer, L. S., and Gorin, M. H. (1942).—" Electrophoresis of Proteins." (Reinhold Publ. Corp. : New York.)

Alexander, A. E. (1947).—Bordeaux conference on surface chemistry. (Supplement to *Research* 1949.)

Alexander, A. E., and Saggers, L. (1948).—*J. Sci. Instrum.* **25** : 374.

Anderson, E. A., and Alberty, R. A. (1948).—*J. Phys. Colloid Chem.* **52** : 1345.

Booth, F. (1948).—*Trans. Faraday Soc.* **44** : 955.

Cumper, C. W. N., and Alexander, A. E. (1950).—*Trans. Faraday Soc.* **46** : 235.

Growney, G. (1941).—*Trans. Faraday Soc.* **37** : 148.

Gutfreund, H. (1948).—*Biochem. J.* **42** : 544.

Hall, J. L. (1941*a*).—*J. Biol. Chem.* **139** : 175.

Hall, J. L. (1941*b*).—*J. Biol. Chem.* **140** : 671.

Moyer, L. S. (1938).—*J. Phys. Chem.* **42** : 71.

Overbeek, J. Th. G. (1950).—" Advances in Colloid Science." Vol. 3. (Interscience Publishers Inc. : New York.)

Powney, J., and Woods, L. J. (1940).—*Trans. Faraday Soc.* **35** : 57.

Wintersteiner, O., and Abramson, H. A. (1933).—*J. Biol. Chem.* **99** : 741.

ADSORPTION AND HYDROGENATION OF GASES ON TRANSITION METALS

By M. E. Winfield*

[*Manuscript received January* 25, 1951]

Summary

Several principles are suggested which determine the location, and to some extent the behaviour, of gas atoms chemisorbed at the surface of a transition metal. With these as basis, and the published experimental evidence, the structure of adsorbed oxygen, hydrogen, ethylene, acetylene, and carbon monoxide are deduced. For a diatomic molecule three kinds of chemisorbed product are distinguished—adsorbed molecule, adsorbed atom, and adsorbed ion. Bond energies, together with solid models of surface and adsorbate, are used to arrive at the probable mechanism of the ethylene hydrogenation on nickel, and of the Fischer-Tropsch hydrocarbon synthesis.

I. Introduction

The chief obstacle to further progress towards an exact knowledge of catalysis is ignorance of the structure of the surface on an atomic scale, and the nature of the bonds which it can form with an approaching gas atom. In the following discussion the experiment is made of building simple models of the surface and the kinds of bonds which may be found there, and from them deriving the probable structure and reactivity of several species of adsorbate.

II. General Principles

In the sections which follow, we assume that the following principles govern the behaviour of metal surfaces towards gas atoms, in addition to those already familiar in the literature on catalysis (e.g. Dowden 1950).

The spacing between atoms, within the outermost layer of a crystal, is required by the inductive effect of the underlying atoms to be approximately the same as that in the bulk of the metal. A model of the surface can therefore be made from equal spheres packed according to the accepted crystallographic constants for the bulk metal, on the usual scale of 1 cm. =1 Å. Both adsorbed atom and the metal atom to which it is covalently bonded are represented by spheres that have been truncated to reduce the interatomic distances to accord with the known covalent radii. Since their coordination number is less, the surface metal atoms may be smaller than those below, but the interatomic distances within the surface plane are assumed to be unaltered.

Attention is confined to crystals which have a face-centred cubic lattice. The (111) plane is close-packed while the (110) is the least dense of the three major planes. Within the crystal are octahedral holes surrounded by the

---

* Division of Industrial Chemistry, C.S.I.R.O., Melbourne.

approximately spherical metal atoms. There is one such hole beneath every second hollow that is seen when looking downwards on to a model of the (111) plane (Fig. 1). Likewise it must be considered that above and including each of the alternate hollows there is an incomplete octahedral hole. We shall assume that a chemisorbed atom tends to be located in one of the octahedral holes.* The location of an adsorbed atom is thus readily decided at any exposed plane, and there is no ambiguity at edges or corners.

The concept is reasonable in that a dissolved gas atom, migrating from within a metal crystal, moves by activated jumps from one interstitial position, that is, octahedral hole, to another (see for example Barrer 1941, p. 153) until when it reaches the surface it is situated in an incomplete octahedral hole. Also it is expected that the atoms of the metal, which already have bonds partly directed (transition metals) towards each other, will preferentially form new bonds in the directions which are free from bonds ; that is towards the centres of either the octahedral or the tetrahedral holes. This is tantamount to saying that two neighbouring bonds are stronger when the bond orbitals do not overlap. Little evidence can be found on this point, but it is highly suggestive that in the structure of the double bond in for example ethylene, the orbitals of the second bond do not coincide in location with the first but spread out considerably beyond them.

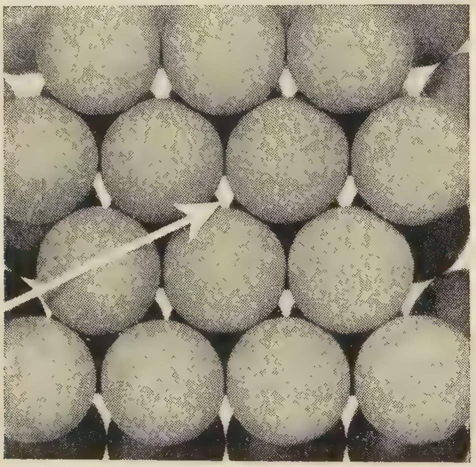


Fig. 1.—Model of nickel crystal showing (111) plane. Of the spaces between the two atomic layers that are visible, those which appear white (arrow) are octahedral holes while those which appear black are tetrahedral.

An important extension of the concept is that when an atom, or to a lesser extent a molecule, is adsorbed by bonds to each of two nearest neighbour metal atoms, the heat of adsorption is less than that calculated from bond energies which are applicable when there is no interference between orbitals.

---

* During migration, or under the influence of a second adsorbate or high temperatures, the atom may be activated to other, less stable, positions.

Chemisorption of a diatomic gas can proceed through a number of intermediates of decreasing potential energy. The gas is finally present in the adsorbed state which has the lowest energy, with the reservation that at low temperatures there may be insufficient thermal energy for activation of steps subsequent to the first or second.

The greater the number of intermediates in the process of adsorption to a given state, the smaller in general will be the activation energy of the overall process (cf. biological processes).

In the case of atoms of about the same electronegativity as the metal, we make the approximation that the adsorbate is held by covalent bonds, rather than the vague " partial bonds " sometimes proposed. Provided that account is taken of the unusual orbitals that are invoked in these bonds, adsorption heats calculated from bond energies have reasonable values, as will be shown later. The bonds from surface to adsorbate are in fact much stronger than has generally been supposed. The deceptively small heat of adsorption of CO, for example, does not mean that it is held by weak bonds, but is the result of loss of some of the exceptionally great resonance energy of the C—O bond in gaseous carbon monoxide.

On a transition metal we consider that all gases are chemisorbed first by covalent bonds. This is followed by a transition to adsorbed ions where the atom is considerably more electronegative than the metal (e.g. O, N, Cl). Because they have greater potential energy, it is generally the adsorbates which are held by bonds predominantly covalent in character that undergo catalytic reaction. Certain adsorbed atoms may have greater mobility also. For their migration along the surface, only the breaking of one of the several bonds by which they are held to the metal appears to be necessary (see end of Section IV).

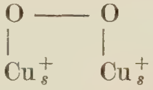
When a gas is very sparsely chemisorbed in a mobile form, the distribution of adsorbed atoms or ions at the surface is not necessarily uniform. On a given crystal plane, we shall assume that uncovered areas exist adjacent to areas completely covered with the most expanded of the possible phases. This phase will consist of a two-dimensional crystal in which the interatomic distances are a simple multiple of those in the underlying metal. When the gas pressure is increased, complete coverage by the first phase may be followed by the appearance of a denser phase, in which the spacing is a simple fraction of that existing previously. The gas crystal will of course become indistinguishable at elevated temperatures.

This concept of discrete phases arises from a compromise between the lattice energy gained (negative $\triangle E$) by an orderly array and the energy lost by interaction of adsorbed atoms with neighbouring sites. There is some experimental support in the slow-electron diffraction studies of Davisson and Germer (1927), and Schlier and Farnsworth (1950).

The simplest kind of metal surface, which we envisage in this discussion, is practically unattainable. Where there are missing or excess metal atoms at a given plane, other crystal planes must be exposed in small amount. But if we consider a sufficiently small area of exposed pure metal, it can be treated as a homogeneous crystal plane. Where the plane is contaminated by adsorbed gas, the following approach can be made: the foreign gas is most likely oxygen. If the metal has been carefully degassed, it can be assumed to be covered at least in part by the most expanded phase of adsorbed oxygen, e.g. for the (100) plane of copper, covered by the double-spaced phase found by Schlier and Farnsworth (1950). For the purpose of hydrogenations, therefore, a given plane is not truly homogeneous. But, just as more oxygen can be adsorbed between the gas atoms of the most expanded phase, so there are small areas which are capable of hydrogenating in much the same way as the uncontaminated metal would. Particularly where this most stable oxygen phase consists of adsorbed ions, the area in the vicinity should be a more active catalyst for certain reactions because the oxygen has increased the electron affinity of the metal.

III. Chemisorption of Oxygen

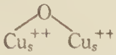
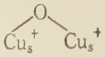
Much can be learnt from examination of the only well-investigated oxygen adsorption—that on cuprous oxide by Garner, Gray, and Stone (1950). We interpret their results to mean that oxygen is first adsorbed as a molecule held by covalent bonds and represented



$$\begin{array}{ccc}\mathrm{O} & \text{——} & \mathrm{O}\\ | & & | \\ \mathrm{Cu}_s^+ & & \mathrm{Cu}_s^+\end{array}$$

where Cu_s^+ is a cuprous ion at the surface.* One electron per atom of oxygen is shared with the metal. The nature of the covalent bond to a metal surface is discussed in Section IV and the ability of copper ions to form such bonds in Section VIII.

Fission next occurs to yield an adsorbate of greater stability, in which two electrons per oxygen atom are shared with metal atoms. The transition is exothermic because two new bonds are formed whose energy together is greater than that of the O—O bond. Some of the adsorbed atoms receive electrons from the cuprous ions, enabling them to become ions essentially the same as oxide oxygen. Their stability is now considerable. The metal oxides are among the most refractory of materials.

If we suppose that unit surface of cuprous oxide has exposed two Cu^+ ions and one $\mathrm{O}^=$ ion, adsorption of a molecule of oxygen yields finally one $\mathrm{O}^=_{\mathrm{ads}}$ at most, since only one electron can be removed from each of the two cuprous ions. The second oxygen atom cannot also drop to the low potential energy of the ion. It must remain held by covalent bonds to one or both of the cupric ions, which are now in effect trivalent. At least for high surface coverage therefore, only one-half of the oxygen adsorbed is in a reactive form, the result found by Garner, Gray, and Stone (1950). The same may be true at low surface coverage. Since  $Cu_s^{++}-O-Cu_s^{++}$ is probably more stable than  $Cu_s^{+}-O-Cu_s^{+}$ the adsorbed oxygen atom is prevented from migrating far from the adsorbed ion.

---

* If the formation of a Cu^+—O bond requires much activation energy (for excitation from the $3d^{10}$ to the $3d^9\,4s$ level), it is necessary to assume that there is always a little Cu^{++} present at a cuprous oxide surface. This is then used to initiate the molecular adsorption of oxygen.

There is no reason to believe that adsorption of oxygen on a clean metal will differ significantly from that described above, except in regard to the relative amounts of each kind of adsorbate. When the surface coverage is very small, for example when a metal has been degassed at high temperatures, we assume that part of the surface is bare and part is covered by the most expanded phase of adsorbed oxygen ions. When we deduce the structure of this phase by the " sphere of influence " method of Davisson and Germer (1927), assuming that the ions occupy the octahedral holes, it corresponds to the least dense phase found experimentally on the (100) plane of copper by Schlier and Farnsworth (1950). Only one site in four is occupied. Following completion of the first phase on admission of a little oxygen, a second of twice the density may appear.

In its most stable state the adsorbed oxygen atom lies not above one metal atom, where it may be found transiently, but between several. On the (111) plane we consider it to be surrounded by three metal atoms to which it forms three resonating bonds of order 2/3, while on the (100) it has four resonating bonds of order 1/2. The corresponding adsorbed ion differs little in situation, but is less readily moved to adsorption positions of higher potential energy, i.e. more remote from the surface. We expect the adsorbed oxygen molecule to exist only transiently, although it possibly occurs at equilibrium at low temperatures when the surface is well covered.

IV. CHEMISORPTION OF HYDROGEN

Hydrogen is not so far different from the metal in electronegativity as to be adsorbed by bonds which are largely ionic in character. Also the hydrogen molecule can form only weak bonds to the metal (see, however, Eucken 1950). Only the adsorption of hydrogen as an atom, therefore, will be considered here.

At a (100) plane, if a hydrogen atom were held by a resonating bond to the four neighbouring nickel atoms, the shortest possible bond length would be 1·76 Å, compared with about 1·5 Å for a single non-resonating bond. A rough calculation, using Badger's rule (Pauling 1940) and Sutherland's (1940) equation, shows that the resonance energy gained would be offset by the weakening due to extension of the bond distance. Hydrogen is therefore too small to form strong bonds when it is placed centrally in an octahedral hole, and tends to occupy the tetrahedral holes when dissolved in the metal. The same should be true of the adsorbed atom, but would require that on the (100) plane it lie between only two metal atoms. Since the latter position is certainly not the lowest surface level available, the most consistent rule for hydrogen adsorption is that the adsorbed atoms, in their most stable state, are situated in the octahedral holes. But on the (100) and (110) planes it is only the corner of a hole that is occupied, so that on all the major crystal planes the hydrogen atom has a bond of order 1/3 to each of three surrounding metal atoms. This largely explains the observed independence of adsorption heat on the type of plane exposed (Beeck 1950*a*).

Except at elevated temperatures the distribution of hydrogen at the surface follows from the principle of discrete phases suggested earlier. At very low pressures the (111) plane for example should be partly bare and partly covered with the most expanded phase that can be drawn by putting circles, which do not quite overlap, around the three nearest-neighbour metal atoms of each adsorbed atom. This is the phase observed by Davisson and Germer (1927), in which the spacing between gas atoms is twice that between metal atoms. With increasing pressure it is succeeded by a phase of twice the density, in which every surface octahedral hole is filled ; that is there is one gas atom for every surface metal atom.

We conclude from the above considerations that the two atoms formed when hydrogen is adsorbed will not remain in neighbouring adsorption sites if the surface coverage is small. At very low pressures one of the atoms must be considered to migrate almost immediately to a site more remote by a factor of at least two. Since, moreover, an adsorbed atom has a short life before desorbing or dissolving, the reactivity of adsorbed hydrogen may be a function of the time elapsing between its admission to the surface, and admission of the second gas (Eucken 1950).

As the potential energy of the metal atoms at the surface is greater than that of their counterparts in the interior, the heat given out when a surface (incomplete) octahedral hole is filled should be greater than that in the centre of the crystal, i.e. for solution of the gas in the metal. Thus, at room temperature or thereabouts, hydrogen will not dissolve appreciably in nickel until most of the surface is covered with adsorbed hydrogen (Beeck, Cole, and Wheeler 1950).

(*a*) *Bond Energies at a Nickel Surface*

Before proceeding to discuss the adsorption of ethylene it is necessary to examine, within the accuracy limits of available bond energies, the energetics of hydrogen adsorption.

The H — H bond energy is taken to be 104 kcal./mole (Skinner 1945). Two figures for the energy of the covalent bond Ni_s — H, between an adsorbed hydrogen atom and a nickel atom which is at the surface of a nickel crystal, can be deduced in addition to the value 53 kcal. suggested by Balandin (1942). The first is based on the value 60 kcal. ascribed by Gaydon (1947) to the dissociation energy of the diatomic molecule NiH. Gaydon assumed an $s — s$ bond, i.e. that the electron from the hydrogen atom occupies the s orbital of the nickel atom. However, this assumption leads to an excessively high value for the Ni_s — H bond energy. We assume instead that a $d — s$ bond, of strength

$2 \cdot 2$ (Pauling 1940), is involved, the electron from the hydrogen atom pairing with an electron in the fourth $3d$ orbital, thus:

$$\left(\begin{array}{cccccc} & & 3d & & & 4s \\ \overline{\uparrow\downarrow} & \overline{\uparrow\downarrow} & \overline{\uparrow\downarrow} & \overline{\bullet} & \overline{\uparrow} & \overline{\uparrow\downarrow} \end{array}\right)$$

where, using Pauling's (1949) notation, ↑ denotes an unpaired electron and a solid circle denotes a bond electron. More than 30 kcal. is gained by forming the $d - s$ bond, compared with the $s - s$ bond suggested by Gaydon (1947). The latter's calculation of the dissociation energy should be unchanged by the alteration, because the difference in energy of the lowest $3d^9\,4s$ and the lowest $3d^8\,4s^2$ nickel levels is negligible. For a similar reason the bond energy is taken as equal to the dissociation energy.

A nickel atom at the surface apparently has six full covalent bonds, just as does its counterpart in the interior. If it had a lesser number, we estimate that the potential energy of a surface atom would be of the order of 50 kcal./g. atom greater than that of one at the centre of the crystal, instead of 15–20 kcal. as appears to be the case (Oriani 1950). Considering a (100) surface plane, one of its constituent atoms has six covalent bonds resonating among eight positions instead of the 12 available for an atom within the crystal. It follows that an adsorbed hydrogen atom cannot be held by an octahedral *dsp* bond (terminology of Pauling 1940). But in the *A* form of nickel (Pauling 1949) there are two unpaired $3d$ electrons, and the first of these we assume to share an orbital with the hydrogen electron.* Thus the $Ni_s - H$ bond can be taken on the simplest view to be an $s - d$ link, of energy 60 kcal.

A second estimation is based on the heat of sublimation of nickel. It is unreliable because no satisfactory method is known for determining the bond energy in a transition metal. To the sublimation heat of 90 kcal. (Dushman 1949) must be added the energy required to excite a nickel atom from the ground state to an imaginary state in which all three $4p$ orbitals are occupied. There is no spectroscopic value for such a level, and we therefore take it to be equal approximately to three times the energy needed to promote one electron in nickel to a $4p$ orbital, namely $3 \times 73 \cdot 6$ kcal. (Bacher and Goudsmit 1932). The energy of one resonating octahedral bond in a nickel crystal is then $(90+220)/3=104$ kcal. Pauling (1949) has not calculated the extent to which this bond is strengthened by resonance, but the alkali metal bonds are shown to be strengthened by a factor of 2 to $2 \cdot 5$. Taking the maximum figure, $2 \cdot 5$, the energy of a non-resonating octahedral bond, of strength $(2 \cdot 9)^2$ (Pauling 1940) is $104/2 \cdot 5=41$ kcal. This leads to an energy 24 kcal. for a hypothetical non-resonating d-bond, of strength $(2 \cdot 2)^2$. After correction for the fact that the bond radius in the metal crystal is greater than the covalent bond radius, this becomes 32 kcal. Badger's rule (Pauling 1940) and Sutherland's (1940) equation for the relation between bond length, force constant, and bond energy, were used in making the correction. According to the geometric mean postulate (Pauling

---

* Couper and Eley (1950) reached a similar conclusion with regard to the adsorption of hydrogen on palladium.

1940), the Ni_s — H bond energy is then $(32\times104)^{\frac{1}{2}}=58$ kcal. The excess binding energy at the surface (Oriani 1950) was neglected because it was estimated to be only about 2 kcal. In view of the doubt regarding the excitation energy, and the many assumptions required by the calculation, the agreement between the two bond energy determinations may be fortuitous.

In later calculations we take the energy of a non-resonating Ni_s — H bond to be 60 kcal. The heat of adsorption of hydrogen held at a sparsely covered surface by non-resonating covalent bonds is then $(2\times60)-104=16$ kcal./mole, compared with 30 kcal. obtained experimentally (Beeck 1950*a*). It was indicated earlier that an adsorbed hydrogen atom is held on the (111) plane between three nickel atoms. By resonance of the Ni_s — H bond between three positions, it is reasonable to expect a strengthening of 5–10 kcal.; that is the energy of a resonating Ni_s — H bond may well be the 67 kcal. required to account for the observed adsorption heat of 30 kcal.

The extraordinary surface mobility of adsorbed hydrogen atoms (Beeck 1950*a*, 1950*b*) is not difficult to understand in terms of the resonating covalent bonds with which we assume they are held. Although the total bond energy is great (67 kcal.), migration does not necessitate breaking more than one of the three resonating bonds. The activation energy required for surface migration is the difference in resonance energy when there are two metal neighbours instead of three, i.e. less than 7 kcal.

V. CHEMISORPTION OF ETHYLENE

Ethylene adsorption on nickel at 23 °C. can result in several kinds of adsorbate, depending on θ, the fraction of surface covered. When θ is small, the ethylene molecule is held at the metal surface by two Ni_s — C covalent bonds. At (110) planes the octahedral hole concept does not distinguish between adsorption on nearest-neighbours, as suggested by Twigg and Rideal (1940), and adsorption on atoms 3·52 Å apart (Fig. 2). The second alternative is favoured because the adsorbed molecule is then more deeply immersed in the metal. If the carbon atoms are rotated with respect to each other so that one of the hydrogen atoms attached to each carbon lies in a hollow between four nickel atoms, little stretching of bonds is required. Choice of the model shown in Figure 2 is in accord with Beeck's (1950*a*) finding that polymerization occurs readily even on (110) planes, at which 90 per cent. or more of the adsorbate is polymer. From an examination of the solid models we can find no mechanism for ready polymerization, at a (110) plane, between adsorbate molecules which have the Twigg and Rideal (1940) structure. At (111) planes this objection does not apply and, moreover, adsorption must occur on nearest-neighbours because there are no metal atoms 3·52 Å apart.

We take the view that adsorbed ethylene has at first a single bond between the carbon atoms, and retains all four hydrogen atoms, contrary to the belief of Beeck (1950*a*). As will be shown later, an adsorbate molecule of this kind has a higher heat of adsorption than any other monomer which can be envisaged. It may well be a general rule that, when a hydrocarbon is adsorbed on nickel, the number of bonds between the two carbons concerned in the adsorption process is reduced by one, while at the same time two C-metal bonds are formed. Subsequent modification of the adsorbate may of course occur.

Ideally, ethylene would be adsorbed in the undissociated form until $\theta = 0 \cdot 5$, forming an expanded phase in which every second pair of adsorption sites is unoccupied. (In practice θ appears to be a little less than 0·5 (Beeck 1950*a*).) A close-packed phase would next appear, with twice the density of the earlier one, and in which the carbon atoms of adjoining molecules are sufficiently close to permit new C — C linkages. Polymerization (Fig. 3) would then commence, for reasons which will become apparent later from consideration of adsorption heats.

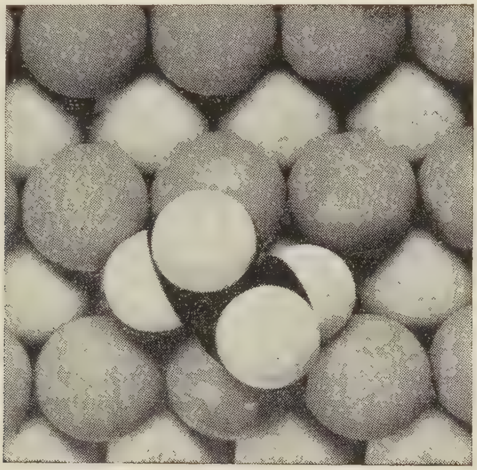


Fig. 2.—Undissociated ethylene adsorbed on the (110) plane of nickel.

Each polymerization step must be accompanied by dissociation of hydrogen atoms. These remain adsorbed only briefly. They react rapidly with gaseous ethylene to yield a molecule of ethane, the latter desorbing almost at once (Beeck 1945). Adsorbed hydrogen is removed so quickly that the fraction of surface occupied by it can be neglected. If hydrogenation of gaseous ethylene requires the 3·52 Å spacing, as implied by Beeck (1950*a*), hydrogen which is liberated at a (111) plane must migrate to a (110) surface before being consumed. However, this also is a rapid process, since hydrogen is mobile on nickel even at −183 °C. (Beeck 1950*b*).

According to our model, rapid reaction ceases when the surface is nearly covered with the close-packed phase. Approximately as many ethylene molecules have now been adsorbed as there are ethane molecules in the gas phase (Beeck 1950*a*). Further ethane production may be dependent on formation of a new type of polymer (Fig. 4), an unsaturated one which is conveniently written $C_nH_{0\cdot 5n}$, neglecting the extra hydrogen atom at each end of the chain. Every alternate $C — Ni_s$ bond is severed, accompanied by loss of a hydrogen atom from the neighbouring carbon atom, so that a double bond is formed, and by repetition a conjugated chain. This process of freeing alternate rows of adsorption sites requires appreciable activation and is only weakly exothermic, possibly even endothermic. It proceeds because of the irreversible removal as ethane of any hydrogen that may be liberated.

Of the polymers which can be constructed with the solid models, only the $C_nH_{0\cdot5n}$ type shown in Figure 4 can fit all three major planes. The simplest polymer to construct from adsorbed ethylene molecules is that shown in Figure 3, but it is confined to the (111) plane. Dimers of the same general composition C_nH_n but different skeletal arrangement can be formed at (110) planes without great distortion of bond angles, but further polymerization appears difficult.

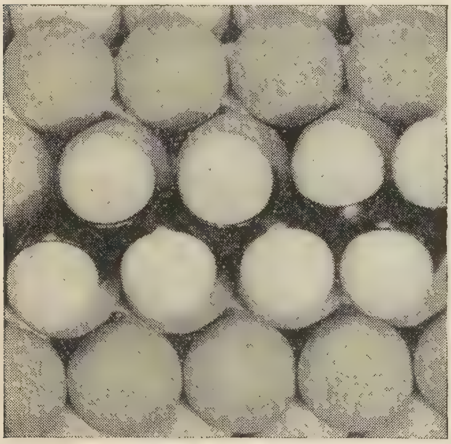


Fig. 3.—Ethylene polymer C_nH_n adsorbed at a (111) plane.

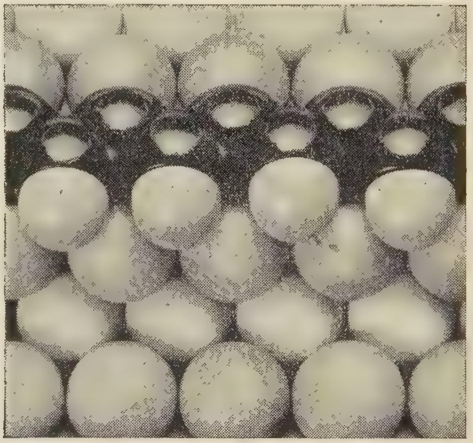


Fig. 4.—Adsorbed ethylene polymer $C_nH_{0\cdot5n}$ at a (110) plane.

It is significant that the product obtained by hydrogenating off the adsorbate formed at high ethylene admissions is the same for oriented and unoriented nickel films (Beeck 1950*a*), which suggests that the nature of the polymer is independent of the crystal plane on which it is held.

Polymers with less than 0·5 H per carbon atom cannot be constructed; nor can branched chain polymers because of the considerable distortion of bond angles that is required.

There is doubt whether definite bond energies can legitimately be ascribed to chemical bonds at metal-gas interfaces, or whether accurate values for such energies can be determined. Even approximate bond energies, however, can be utilized to obtain adsorption heats which indicate the relative probability, and order of formation, of the members of a series of hypothetical adsorbates on a given metal. A similar procedure has been employed by Balandin (1942).

Skinner's (1945) values for carbon-bond energies, assuming the heat of sublimation of carbon to be 170 kcal., are taken as sufficiently accurate for the present computations. Table 1 shows the bond energies which we have employed.

The Ni_s — C bond energy was calculated from that for Ni_s — H, using the geometric mean postulate and assuming a non-resonating bond between carbon and metal. Sufficient thermal data for an independent computation by the method of Long and Norrish (1949) are not yet available.

TABLE 1

BOND ENERGIES USED IN CALCULATING REACTION HEATS

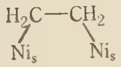
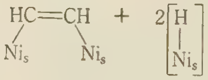
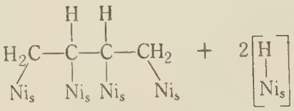
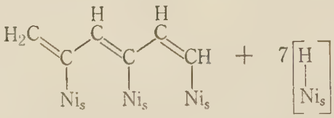
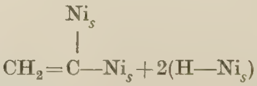
Bond	Compound	Energy/ Mole (kcal.)
C — C	Ethane and higher homologues . .	84
C = C	Ethylene	131
C ≡ C	Acetylene	186
C — H	Ethane	98
C — H	Ethylene	101
C — H	Acetylene	102
H — H	Hydrogen	104
Ni_s — H	Adsorbed hydrogen	60
Ni_s — H	Adsorbed hydrogen (resonating bond) . .	67
Ni_s — C	Adsorbed hydrocarbon (tetrahedral carbon atom)	54

For easy comparison, the heats of adsorption given in Table 2 are calculated on the basis of one mole of ethylene in each case, and for a very small value of θ so that the decline in adsorption heat with surface coverage need not be considered. Providing that a reaction is geometrically possible, the energies given in the table indicate to a first approximation the likelihood of its occurrence relative to other conceivable reactions. It will be noticed that the energies are in harmony with the dimerization of ethylene adsorbed on nickel, and its further polymerization. Also that the bond between hydrogen and carbon is so strong that it generally will not be broken, except when the energy loss can be compensated by formation of a C — C in addition to a C — Ni_s bond, the latter being of only about half the strength of a C — H link. Thus dissociation of hydrogen possibly does not occur except accompanied by polymerization. A small concentration of dissociated monomer may be permissible, however, because of the irreversibility of hydrogen removal by gaseous ethylene. This uni-directional process of removal of hydrogen, as ethane, may well be more important in explaining why polymerization occurs than are the energy considerations.

The adsorption of ethylene on other transition metals will be analogous to that on nickel. Differences which are observable experimentally should be explainable in terms of the ease of polymerization on a given metal, and the relative values for the H-metal and C-metal bond energies. When comparing the latter quantities, it is not sufficient to consider the sublimation heats of the metals concerned. In spite of a higher heat of sublimation, rhodium for example forms a weaker bond with hydrogen than nickel does. The excitation energy referred to earlier must also be included, and this energy is practically uninvestigated within the transition metals.

TABLE 2

ESTIMATED HEATS OF ADSORPTION FOR ALTERNATIVE ETHYLENE MONOMERS AND POLYMERS

Adsorbate	Structure	Heat of Adsorption/ Mole of C_2H_4 (kcal.)
A. Undissociated monomer ..		49
B. Dissociated monomer ..		40*
C. Saturated dimer		60
D. Saturated trimer		63·7
E. Saturated tetramer		65·5
F. Saturated pentamer		66·6
G. Unsaturated trimer		31†
H. Unsaturated pentamer		33†
Rejected Adsorbate Monomers		
I. Undissociated monomer, 1-point adsorption	$CH_3-CH=Ni_s$	28**±10
J. Partly dissociated monomer, 1-point adsorption	$CH_2=CH-Ni_s+H-Ni_s$	20
K. Dissociated monomer, 1-point adsorption	$CH_2=C=Ni_s+2(H-Ni_s)$	16**±10
L. Dissociated monomer, 2-point adsorption	 $\begin{array}{c}Ni_s\\ \vert\\ CH_2=C-Ni_s+2(H-Ni_s)\end{array}$	<40

---

* Possibly 7 kcal. should be added as the approximate resonance energy for a C=C bond at a metal surface.

† Possibly 5 kcal. resonance energy should be added.

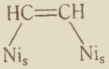
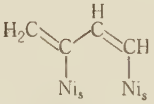
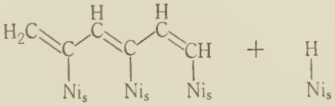
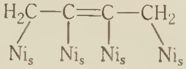
** Assuming 84±10 kcal. for the $Ni_s=C$ bond energy.

VI. Chemisorption of Acetylene

There is not sufficient evidence from experiment to determine the details of acetylene adsorption on nickel. The most consistent assumption that can be made is that the first step is analogous to that in ethylene adsorption and yields the adsorbate *A* in Table 3. *A* is identical with the dissociated ethylene monomer (*B* in Table 2), but the latter constitutes only a small proportion of the mixture of products which is adsorbed ethylene. Acetylene does not polymerize on nickel at room temperature in a similar manner to ethylene because the dimer *D* in Table 3 is most unlikely to occur. Since formation of the dimer *B* is energetically improbable, polymers may be absent altogether. Self-hydrogenation is energetically unsound.

Table 3

CALCULATED ADSORPTION HEATS FOR ACETYLENE

Adsorbate	Structure	Heat of Adsorption/ Mole of C_2H_2 (kcal.)
A. Undissociated monomer . .		51*
B. Unsaturated dimer, 2-point adsorption		39†
C. Unsaturated trimer		42†
D. Dimer** with 4-point adsorption		63·5
E. Undissociated monomer, upright adsorption	$H_2C{=}C{=}Ni_s$	27 ± 10
F. Partly dissociated monomer, 1-point upright adsorption	$HC{\equiv}C{-}Ni_s + H{-}Ni_s$	19

* Plus 7 kcal. resonance energy.

† Plus 5 kcal. resonance energy.

** Geometrically impossible on a model surface.

VII. HYDROGENATION OF ETHYLENE

We shall assume that the fast reaction responsible for most of the ethane produced in the presence of hydrogen follows the same mechanism as the fast self-hydrogenation reaction which has been investigated by Beeck (1950*a*). It is convenient to describe the reaction on the (110) plane of nickel under conditions where the surface is little more than half-covered with adsorbed ethylene. Part consists of monomer spaced in such a way that every alternate pair of nickel atoms 3·52 Å apart is free. Ethane forms when an ethylene molecule from the gas phase collides with one of these pairs of nickel atoms, at an instant when they are bonded to hydrogen. Beeck apparently supposes that each of the pair of atoms holds a hydrogen atom, hydrogen which has dissociated from neighbouring adsorbed ethylene. We prefer to say that the hydrogen is not vertically above the nickel atom, but may be in an adjacent octahedral hole so that it has only a bond of order 1/3 with the nickel atom with which the ethylene collides.

In competition with the direct hydrogenation there is a reaction in which ethylene, following collision, is adsorbed on the two nickel atoms. This presumably occurs when no adsorbed hydrogen is in the vicinity at the time of collision. It may be followed by polymerization with the adsorbed ethylene on either side unless quickly removed by hydrogen migrating to the site.

When both ethylene and hydrogen are admitted to a nickel surface at room temperature, the observed rate of ethane production has been shown by Beeck (1945, 1950*a*) to be largely the rate of the direct hydrogenation occurring on that proportion of the exposed (110) planes that is not covered with adsorbed ethylene. Hydrogenation at the (111) or (100) plane is at least five times as slow (Beeck 1950*a*) and will therefore be neglected. Under steady state conditions the area that is effective is maintained by a reaction in which ethylene, present at the surface as monomer, is removed by hydrogenation to ethane. We believe that removal of polymer is too slow to be important in this respect, and polymerized ethylene can be treated as a relatively permanent poison of a large part of the surface.

(*a*) *Hydrogen Migration during Catalysis*

As the surface is progressively covered with ethylene, surface migration of hydrogen will be increasingly hindered and may ultimately be less than subsurface migration at a depth of about 3 Å. There is reason to believe that the activation energy of the latter process is quite small, since Beeck (1950*b*) has found that solution of hydrogen in nickel takes place at 80 °K. or even lower.

It is a significant characteristic of many catalytic systems that one of the reactants or products dissolves in the catalyst (see for example Wagner and Hauffe 1945). In the ethylene hydrogenation on nickel, we suggest that the clean-up of adsorbed ethylene is achieved by hydrogen atoms which emerge at the surface very close to carbon atoms that are bound to the metal. If the dissolved hydrogen atom is looked upon as diffusing from one tetrahedral hole to another in a series of activated jumps, it will eventually reach a tetrahedral hole which is at the surface. If there is a carbon atom attached to one of the three nickel atoms which surround it, the hydrogen atom may attach itself to the carbon rather than diffuse to below the surface again. Thus the effective hydrogen concentration may be appreciably greater than would be anticipated at a surface which is largely covered with hydrocarbon.

Couper and Eley (1950) have found that hydrogen dissolved in palladium poisons the metal for the parahydrogen conversion. It cannot, however, be concluded that the effect is a poisoning for all hydrogenations. For example, solution of hydrogen in nickel may increase the activation energy for hydrogen adsorption, yet benefit the ethylene hydrogenation by an equal or greater increase in the activation energy for ethylene adsorption (the latter being an undesirable reaction). That the activation energy for adsorption of ethylene is greater than for hydrogen is shown by the minimum temperatures at which each process can occur (Beeck 1950*a*, 1950*b*).

(*b*) *On the Significance of Lattice Parameter*

Although the lattice parameter of a metal is in part determined by electronic configuration, and the latter is undoubtedly of great importance in determining catalytic activity (Dowden 1950 ; Couper and Eley 1950), the lattice spacing itself is critical in many catalytic reactions. Its influence is greatest when two atoms of an adsorbate are bonded to each other as well as to the metal, bond energy being steeply dependent on bond length. If a bond between two carbon atoms requires stretching for adsorption, the adsorption is so much the more difficult, but once achieved it results in a system of higher potential energy. Hydrogenation is therefore easier ; that is, the lower heat of adsorption when the lattice parameter is increased is compensated by the smaller energy of activation required for the surface hydrogenation. The observable activation energy of the overall reaction, which includes the difference of these two energies, may therefore be independent of lattice dimension. In contrast, the corresponding frequency factor may change by a substantial amount, as found by Beeck (1945) for the ethylene hydrogenation.

When ethylene is the reactant, lattice dimension has an additional importance in determining the distance between the carbon atoms of adjacent adsorbed molecules, and therefore the strength of the bonds formed during polymerization. On rhodium this distance is great enough to hinder polymer formation. The product, furthermore, has a lower total bond energy than on nickel, and requires correspondingly less activation energy for its removal.

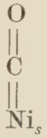
Already on rhodium, when ascending the transition series arranged in order of lattice parameter, polymerization at 23 °C. occurs to the extent of only a few per cent. (Beeck 1950*a*), while the lattice spacing is ideal for the direct hydrogenation of gaseous ethylene. Metals with longer spacing are less suitable in the latter respect, while at the same time being poisoned negligibly less by polymer.

VIII. Chemisorption of Carbon Monoxide

Below 100 °C. carbon monoxide must be adsorbed on transition metals by bonds from carbon to metal.* There is no evidence of interaction between adsorbed molecules leading to production of other adsorbate species, for example, polymers. At higher temperatures some oxygen dissociates to leave surface carbide (Craxford 1947). It is the irreversible removal as CO_2 of any oxygen which dissociates that accounts for carbide being found in detectable amount.

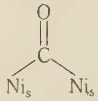
Since adsorption of CO, like that of hydrogen, takes place rapidly even at extremely low temperatures (Beeck 1950*a*, 1950*b*), any formula for the adsorbate must imply an adsorptive mechanism which requires little activation, and bonds in which electrons are shared with the metal. Extraction of electrons to permit ionic bonding is now generally recognized to be a process needing appreciable activation energy.

Localization of two electrons, from each surface nickel atom that forms a bond with CO, is necessitated by the conventional structure:



$O=C=Ni_s$

The localization can be reduced by participation of two nickel atoms, giving:



and reduced further by participation of three or four nickel atoms. When the heats of adsorption of the above two forms are approximately determined from the bond energies, taking into account the bond-weakening which is assumed when adsorption occurs on two nearest-neighbour metal atoms, the two adsorbates are found to be of about equal stability.

If the carbon atom is instead situated in the hollow between four nickel atoms, at a (100) plane for example, it is doubtful whether it can, in view of the existing bond to oxygen, form four resonating bonds of order 1/2. Such a configuration would, however, account for the small activation energy for adsorption, as well as considerably increasing the adsorption heat above the value of about 20 kcal. estimated for the two structures given above. An accurate experimental determination of the heat of adsorption of CO on clean nickel may be sufficient to establish the structure of the adsorbate.

The little evidence that is available regarding CO adsorption on metals can be interpreted to mean that this gas is not chemisorbed by copper (van Itterbeek and van Dingenen 1941), except when it contains an impurity which has extracted electrons from the copper, making it in effect a transition metal. The same appears to be true for hydrogen adsorption (Beeck 1950*a*). This lends support to our belief that in both adsorptions the bonding is achieved by an electron from a gas atom sharing an orbital with an unpaired *d*-electron of a metal atom. If CO were held at the surface by a predominantly ionic bond, adsorption would require that the gas molecule extract electrons from a metal atom—a process which is possible not only on the transition metals but on copper also, as demonstrated by the adsorption of oxygen.

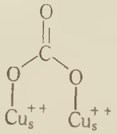
---

* Adsorption in which both carbon and oxygen form a bond with the metal cannot occur because of the unfavourable bond angle introduced by a $C = Ni$ link. One-point adsorption through the oxygen atom is also unlikely. It necessitates a monovalent carbon atom. Adsorption as the ion $C^- = O$ would resemble oxygen adsorption and therefore be relatively slow.

(*a*) *Oxygenated and Oxide Surfaces*

In the Fischer-Tropsch reaction the metal surface is partly covered with oxygen. It is therefore important to consider the adsorption of CO on surfaces where both metal and oxygen are available for bonding with adsorbate. The results of Garner, Gray, and Stone (1950) with cuprous oxide surfaces suggest that CO will be first adsorbed as it is on metallic nickel. Copper ions can be treated as nickel atoms for the purpose of adsorption, the presence of oxygen having created unpaired electrons with which electrons from the adsorbate can share a 3*d* orbital.

When there is adsorbed oxygen present (that is, oxygen atoms held by covalent bonds to copper ions) as well as the oxide oxygen, a transition can occur to a second kind of adsorbed CO (Garner, Gray, and Stone 1950), which probably can be written:



When desorption occurs, CO_2 rather than CO appears in the gas phase. In the Fischer-Tropsch reaction therefore, some of the CO_2 produced may be the result of direct oxidation of CO, rather than the product of the water-gas shift reaction.

We conclude that on a Fischer-Tropsch catalyst in the presence of hydrogen, most of the CO, and certainly the CO which will be hydrogenated, is held by C-metal bonds.

IX. Hydrogenation of Carbon Monoxide to Hydrocarbons

Beeck's (1945, 1950*a*, 1950*b*) and Beeck, Cole, and Wheeler's (1950) detailed investigations of the behaviour on nickel surfaces of ethylene, ethane, and hydrogen provide much information which can be applied to the hydrogenation of CO.

Up to the point of production of methylene radicals, the Fischer-Tropsch reaction has been explained by a mechanism suggested by Craxford (1947), and supported by other investigators. Recently Griffiths (1950) has indicated the way in which subsequent polymerization to large molecules may occur. An alternative is that once a C_2 hydrocarbon has been produced, further increase in chain-length follows approximately the course which we have described above for ethylene polymerization on nickel; that is, the growing hydrocarbon is held by bonds from the metal to every carbon atom, or every second atom. If the product is later readsorbed, it will be held by only two C-metal bonds, and will then resemble Griffith's polymer.

Certain curious features of the Fischer-Tropsch process can now be explained. It has long been known (cf. Storch *et al.* 1948) that ethylene, added to the reactants, enters into the synthesis. Yet no ethylene can be detected as an intermediate in the CO hydrogenation. The results obtained by Beeck (1950*a*) suggest that adsorbed ethylene, in one of its several forms, may indeed be an intermediate ; while at the same time showing clearly that no ethylene would be detectable, because its adsorption on nickel is irreversible. Once the ethylene is adsorbed, it can be removed only as ethane or a larger molecule.

Under certain conditions Storch *et al.* (1948) actually obtained a synthesis product which included appreciable ethylene and propylene, especially when an iron catalyst was used. These could not have been the result of desorption of olefine held at a metal surface, so that their origin must be in a dehydration process.

A typical Fischer-Tropsch catalyst contains, besides a transition metal, one or more oxides such as thoria and a carrier, generally kieselguhr. It is commonly believed that the oxide is required to modify the chain-length of the reaction product. If Craxford (1947) is correct in claiming adsorbed methylene as an intermediate, the oxide has a further function in facilitating dimerization, because Beeck (1950*a*) has found that methylene radicals at a nickel surface do not readily polymerize. The most difficult step to elucidate in the CO hydrogenation to higher hydrocarbons is the formation of the first C — C bond, and it is this step which has received least attention.

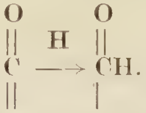
(*a*) *Suggested Mechanism for the Fischer-Tropsch Reaction*

Methylene is not necessarily an intermediate in the synthesis of higher hydrocarbons, since the detectable methylene could have been produced by readsorption of ethane. This yields two adsorbed methyl radicals, which dissociate to some extent to methylene (Beeck 1950*a*), or it could be an intermediate only in the synthesis of methane.

Until there is more experimental evidence for the mechanisms that are based on methylene as an intermediate, it may be stimulating to consider an alternative scheme which arises naturally from the outlook on surface processes that we have presented.

We make the simplest assumption regarding the nature of the surface on which the chief reactions proceed, namely, that it is a metal surface, nickel for convenience. This is consistent with the fact that Fischer-Tropsch catalysts are based on a transition metal, such as iron, cobalt, or nickel. If the important reaction takes place on an oxide surface, then the transition metal should be replaceable by copper, which in the presence of oxygen could also adsorb CO and hydrogen.

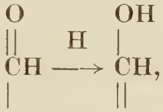
On catalysts, which yield mainly methane, the following reaction can be assumed to occur, following adsorption of the two reactants:



$$\begin{array}{c}\mathrm{O}\\ \|\\ \mathrm{C}\\ \|\end{array}\xrightarrow{\mathrm{H}}\begin{array}{c}\mathrm{O}\\ \|\\ \mathrm{CH.}\\ |\end{array}\quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\quad(1)$$

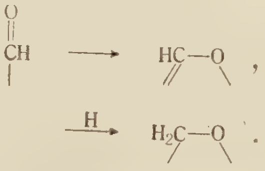
This step, in which the first hydrogen atom is added to carbon, we estimate from the appropriate bond energies (Skinner 1945) to be more probable than the alternative, in which oxygen is attacked first. When the reactants are taken to be adsorbed CO and adsorbed hydrogen, the overall energy change is nearly zero. The reaction heats used in this section are averages of several calculations using alternative plausible energies for the $C = O$ and $C = Ni_s$ bonds in adsorbed carbon monoxide, and its hydrogenation products.

There are now two possibilities for the second reaction. If the oxygen atom is next attacked:



$$\begin{array}{c}\mathrm{O}\\ \|\\ \mathrm{CH}\\ |\end{array}\xrightarrow{\mathrm{H}}\begin{array}{c}\mathrm{OH}\\ |\\ \mathrm{CH,}\\ \|\end{array}\quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\quad(2)$$

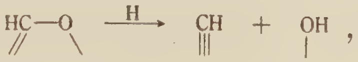
then hydrogenation to $=CH_2$, and finally methane, cannot be expected. It is well known that in the Clemmensen reduction of a carbonyl group complete reduction is often possible when $=CHOH$ can be shown not to be an intermediate (see for example Farkas and Farkas 1939). The second reaction is therefore more likely to be one in which oxygen forms a bond with a neighbouring nickel atom, followed by further hydrogenation of the carbon atom:



.... (3)

.... (4)

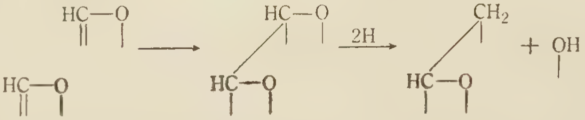
Use of the bond energies does not permit a decisive choice between reaction (4) and the following:



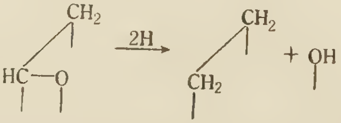
.... (5)

but both should lead ultimately to methane. We estimate that the heat input required for reactions (2) and (3) is the same (20–30 kcal.) within 1 kcal. It is appreciably larger than for any of the other reactions which we propose, so that either (2) or (3) may prove to be the slow step in the Fischer-Tropsch process.

On an efficient polymerization catalyst the same series of reactions, from CO to CH_4, will tend to take place, but with the difference that at some point before methane is reached, the main volume of hydrogenation is diverted along another path. Either $\underset{\|}{\mathrm{CH}}-\underset{|}{\mathrm{O}}$ or $=CHOH$ therefore initiates the reaction which leads to the higher hydrocarbons. The first of these is preferred because the polymerization is known to be favoured by conditions in which there is a deficiency of hydrogen. A model of surface and adsorbate can be used to show that polymerization is more feasible geometrically when the adsorbate molecule is held by bonds from two of its atoms. Not only is it expected therefore that further growth will take place readily once dimerization has occurred, but that dimerization of the C_1 carbon compound will be facilitated if it is adsorbed by way of the oxygen as well as the carbon atom. We therefore write as the most probable dimerization reaction

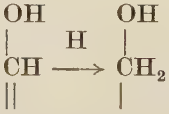


This may be followed by hydrogenation to adsorbed ethylene :



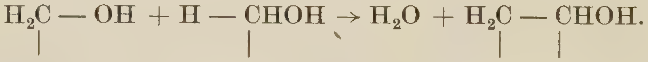
which can either polymerize or be hydrogenated to gaseous ethane. Chain growth ceases when the molecule is hydrogenated to saturated hydrocarbon, or more rarely saturated alcohol. Further along the catalyst bed the alcohol may be readsorbed on an oxide surface, and dehydrated to olefine, or the hydrocarbon may be readsorbed by fission, and finally hydrogenated off as two smaller hydrocarbon molecules.

An alternative mechanism must be considered if reaction (**2**) does actually occur. Since (**2**) can be followed by :



$$\begin{array}{c} OH \\ | \\ CH \\ \| \end{array} \xrightarrow{H} \begin{array}{c} OH \\ | \\ CH_2 \\ | \end{array}$$

which requires approximately zero energy input, $-CH_2OH$ is now the monomer. When there is oxide surface in the vicinity, it is reasonable to suppose that this can condense according to the following reaction :



$$\underset{|}{H_2C} - OH + H - \underset{|}{C}HOH \rightarrow H_2O + \underset{|}{H_2C} - \underset{|}{C}HOH.$$

Chain growth is halted on the resumption of hydrogenation, which leads to an alcohol or hydrocarbon.

X. Acknowledgments

This paper is largely based on the published investigations of gas reactions at metal surfaces carried out by the late Dr. Otto Beeck and his collaborators. Many discussions with and helpful suggestions by Dr. A. L. G. Rees, Division of Industrial Chemistry, C.S.I.R.O., are gratefully acknowledged.

XI. REFERENCES

BACHER, R. F., and GOUDSMIT, S. (1932).—"Atomic Energy States." (McGraw-Hill Book Co.: New York.)

BALANDIN, A. A. (1942).—*Acta Physicochim. U.R.S.S.* **17**: 73-81.

BARRER, R. M. (1941).—"Diffusion in and through Solids." (Cambridge Univ. Press: Cambridge.)

BEECK, O. (1945).—*Rev. Mod. Phys.* **17**: 61-71.

BEECK, O. (1950*a*).—*Discuss. Faraday Soc.* **8**: 118-28.

BEECK, O. (1950*b*).—"Advances in Catalysis." Vol. 2. (Academic Press: New York.)

BEECK, O., COLE, W. A., and WHEELER, A. (1950).—*Discuss. Faraday Soc.* **8**: 314-21.

COUPER, A., and ELEY, D. D. (1950).—*Discuss. Faraday Soc.* **8**: 172-84.

CRAXFORD, S. R. (1947).—*J. Soc. Chem. Ind. Lond.* **66**: 440-4.

DAVISSON, C., and GERMER, L. H. (1927).—*Phys. Rev.* **30**: 705-70.

DOWDEN, D. A. (1950).—*J. Chem. Soc.* **1950**: 242-65.

DUSHMAN, S. (1949).—"Scientific Foundations of Vacuum Technique." (John Wiley & Sons Inc.: New York.)

EUCKEN, A. (1950).—*Discuss. Faraday Soc.* **8**: 128-34.

FARKAS, A., and FARKAS, L. (1939).—*J. Amer. Chem. Soc.* **61**: 1336-41.

GARNER, W. E., GRAY, T. J., and STONE, F. S. (1950).—*Discuss. Faraday Soc.* **8**: 246-50.

GAYDON, A. G. (1947).— "Dissociation Energies." (Chapman and Hall Ltd.: London.)

ITTERBEEK, A. VAN, and DINGENEN, W. VAN (1941).—*Z. Phys. Chem.* B **50**: 341-60.

LONG, L. H., and NORRISH, R. G. W. (1949).—*Philos. Trans.* A **241**: 587-617.

ORIANI, R. A. (1950).—*J. Chem. Phys.* **18**: 575-8.

PAULING, L. (1940).—"The Nature of the Chemical Bond." (Cornell Univ. Press: New York.)

PAULING, L. (1949).—*Proc. Roy. Soc.* A **196**: 343-62.

SCHLIER, R., and FARNSWORTH, H. E. (1950).—*Phys. Rev.* **78**: 316.

SKINNER, H. A. (1945).—*Trans. Faraday Soc.* **41**: 645-62.

STORCH, H. H., ANDERSON, R. B., HOFER, L. J. E., HAWK, C. O., ANDERSON, H. C., and GOLUMBIC, N. (1948).—Tech. Pap. Bur. Min. Wash. No. 709.

SUTHERLAND, G.B. B. M. (1940).—*J. Chem. Phys.* **8**: 161-4.

TWIGG, G. H., and RIDEAL, E. K. (1940).—*Trans. Faraday Soc.* **36**: 533-7.

WAGNER, C., and HAUFFE, K. (1939).—*Z. Elektrochem.* **45**; 409-25.

THE CRYSTAL STRUCTURE OF POTASSIUM HEXAFLUOROPLATINATE (IV)

By D. P. Mellor* and N. C. Stephenson*

[*Manuscript received April 3, 1951*]

Summary

The crystal structure of potassium hexafluoroplatinate (IV) has been determined. The structure is based on a hexagonal unit cell whose dimensions are $a=5 \cdot 76$ Å, $c=4 \cdot 64$ Å. There is one molecule per cell and the space group is D_{3d}^3—$C\bar{3}m$. The Pt–F distance in the slightly irregular octahedral fluoro complex is $1 \cdot 91$ Å.

I. Introduction

Only two hexafluoro complexes of the platinum metals appear to have been described, namely those of platinum itself and iridium (Schlesinger and Tapley 1924). The former compound (K_2PtF_6) has recently been reinvestigated by Emeleus and Sharpe (1950) who devised what appears to be a new and more satisfactory method of preparing it. The crystals (mostly in the form of small, lemon-yellow, thin, hexagonal plates) used in this investigation were very kindly supplied to us by Professor H. J. Emeleus and Dr. A. G. Sharpe.

II. Determination of Unit Cell and Space Group

Powder photographs taken with a 19 cm. camera and using Cu$K\alpha$ radiation ($\lambda=1 \cdot 54$ Å) could be indexed on the basis of a hexagonal unit cell with dimensions :

$$a=5 \cdot 76 \text{ Å},$$
$$c=4 \cdot 64 \text{ Å}.$$

The powder lines are listed in Table 1.

Laue photographs of thin hexagonal plates taken with general radiation from a copper target and with the beam perpendicular to the hexagonal face showed an axis of sixfold symmetry. With the beam parallel to the a-axis, two mutually perpendicular planes of symmetry were found ; two similar planes of symmetry were found with the beam perpendicular to both a and c. These photographs suggest that the crystals belong to the class D_{6h}.

An oscillation photograph of a crystal accurately set about the axis by the method of Weisz and Cole (1948) confirmed the identity period of $5 \cdot 76$ Å. Photographs taken with the crystal rotated about the common perpendicular to two opposite sides of the base of the hexagonal cell showed an identity period

---

* Department of Chemistry, University of Sydney.

9·97 Å which is approximately $\sqrt{3}a$. Weissenberg photographs about the a-axis showed no systematic absences in the $0kl$ spectra. A complete set of oscillation photographs about the c-axis showed no systematic absences of $hki0$. They also gave an identity period of 4·64 Å. A test for pyro-electricity gave no indication that a centre of symmetry was absent.

TABLE 1

LINES IN POWDER PHOTOGRAPH OF K$_2$PtF$_6$

$\sqrt{I}$	Sin² θ$_{obs.}$	Sin² θ$_{calc.}$	hkl	$\sqrt{I}$	Sin² θ$_{obs.}$	Sin² θ$_{calc.}$	hkl
s	0·0235	0·0238	100	*vw*	0·3082	0·3094	310
ms	0·0275	0·0276	001	*vw*	0·3127	0·3132	221
vs	0·0513	0·0514	101	*ms*	0·3199	0·3201	113
ms	0·0714	0·0714	110	*ms*	0·3438	0·3439	203
s	0·1226	0·1228	201	*w*	0·4074	0·4084	401
s	0·1347	0·1344	102	*vw*	0·4146	0·4153	213
w	0·1660	0·1666	210	*w*	0·4198	0·4199	312
w	0·1822	0·1820	112	*w*+	0·4788	0·4798	321
ms	0·1944	0·1942	211	*w*+	0·4904	0·4914	402
ms	0·2061	0·2058	202	*w*+	0·4983	0·4998	410
w	0·2141	0·2142	300	*s*	0·6091	0·6088	214
w	0·2415	0·2418	301	*vw*	0·7045	0·7056	502
vw	0·2500	0·2488	003	*w*	0·7270	0·7278	224
vw	0·2735	0·2725	103	*s*	0·7505	0·7516	314
ms	0·2772	0·2772	212	*vw*	0·7623	0·7624	115
ms	0·2848	0·2856	220	*vw*	0·7852	0·7862	205

The measured density of the crystal (ρ=3·32 g./cc.) showed that there is one molecule per cell. The fact that no systematic absences of reflections were observed showed that the space group could be any one of the following :

$$D_{6h}^1,\ D_6^1,\ D_{3d}^3, \text{ or } D_{3d}^1.$$

The symmetry of the Laue photographs was consistent with D_{6h}^1. Since there is only one molecule per unit cell, D_{6h}^1would require a planar hexagonal PtF$_6^=$ complex, an arrangement so highly improbable and incompatible with the cell dimensions that this space group is excluded. D_{3d}^1 would place the K atoms at $\frac{1}{3}, \frac{2}{3}, \frac{1}{2}$ and $\frac{2}{3}, \frac{1}{3}, \frac{1}{2}$; there is no (002) reflection but there are moderately strong (003), (004), and (005) reflections. This would seem to rule out D_{3d}^1 ; the negative result of the test for pyro-electricity is used to exclude D_6^1. The crystal was therefore assigned to D_{3d}^3 ($C\bar{3}m$). The cell dimensions of K$_2$PtF$_6$ (a=5·76 ; c=4·64) are close to those of K$_2$GeF$_6$ (a=5·62 ; c=4·65) (Hoard and Vincent 1939), and there is no doubt that the two substances are isomorphous. The unusually high symmetry of the Laue photograph taken with the beam parallel to the c-axis which appears to be sixfold (instead of threefold as required by D_{3d}^3) is believed to be due to contact twinning of the crystals on the basal plane.

The coordinates of the special positions of D_{3d}^3 are as follows:

(1) (*a*) 0,0,0. (*b*) $0,0,\frac{1}{2}$.

(2) (*c*) $0,0,z$; $0,0,\bar{z}$. (*d*) $\frac{1}{3},\frac{2}{3},z$; $\frac{2}{3},\frac{1}{3},\bar{z}$.

(3) $x,\bar{x},z$; $x,2x,z$; $2x,\bar{x},z$;
$\bar{x},x,\bar{z}$; $\bar{x},2\bar{x}\bar{z},$; $2x,x,\bar{z}$.

III. DETERMINATION OF ATOMIC PARAMETERS

Consideration of the special and general atomic coordinates shows that an electron density projection along the *b*-axis will involve all the general parameters. The superimposition of two fluorine atoms in this projection will help to intensify an otherwise weak peak. The projection is shown in Figure 1.

TABLE 2

CALCULATED AND OBSERVED STRUCTURE FACTORS FOR K_2PtF_6

hkl	$F_{obs.}$	$F_{calc.}$	*hkl*	$F_{obs.}$	$F_{calc.}$	*hkl*	$F_{obs.}$	$F_{calc.}$
100	14	21	301	14	12	601	13	13
200	12	11	302	16	11	602	5	5
300	20	20	303	20	18	603	10	10
400	12	9	304	16	16	604	11	11
500	13	9	305	12	8	605	11	11
600	17	19	306	8	8			
700	13	12	307	6	13	701	12	12
800	10	8				702	5	9
			401	25	20	703	10	9
101	28	20	402	18	18	704	6	11
102	15	19	403	9	8			
103	12	15	404	9	8	001	16	21
104	14	13	405	11	15	002	3	2
105	14	13	406	10	13	003	16	17
106	12	12	307	6	6	004	20	25
107	6	9				005	12	11
			501	13	14	006	7	4
201	6	7	502	16	16	007	8	12
202	25	25	503	13	10	008	6	14
203	22	19	504	12	8			
204	10	8	505	5	12			
205	9	8	506	6	12			
206	11	14	507	6	7			
207	10	13						

The *h0l* intensity data were obtained by using differently timed Weissenberg photographs taken about the zero level of a *b*-axis rotation. A pack of four films was used in each exposure with an interfilm absorption factor of 7/3. Owing to very high absorption of Cu*K*α radiation by the crystal, it was found necessary to use Mo*K*α in order to obtain a reasonably clearly defined projection. Computation of the Fourier projection was carried out by means of Beevers-

Lipson strips. One-half only of the side *ac* was projected and because of the small cell size each side was divided into thirtieths.

If the Pt atoms are placed at 0,0,0, three parameters remain to be determined, namely x_F, z_F, and z_K. Since the Pt atom whose scattering power far outweighs that of all the other atoms is located at a centre of symmetry 0,0,0, the phase angles of all the reflections are positive.

The following parameters were obtained from the electron density projection:

$$x_F = 0\cdot150,$$
$$z_F = 0\cdot250,$$
$$z_K = 0\cdot740.$$

An approximation for x_F was also found by considering reflections of the type $hk0$. The structure factor for reflections of this type involves only the parameter x_F. By varying the values of x_F, plots of $\sqrt{I}$ against x_F were made for 15 reflections. Then by applying the method of intensity inequalities and using particularly the relation between the reflections

$$4\bar{3}\bar{1}0 > 5\bar{3}\bar{2}0$$
$$43\bar{7}0 \geqslant 6\bar{3}\bar{3}0$$
$$\bar{1}4\bar{3}0 > 6\bar{1}\bar{5}0$$

a value of 0·154 was obtained for x_F. Structure factors calculated from the above parameters are compared with the observed values in Table 2.

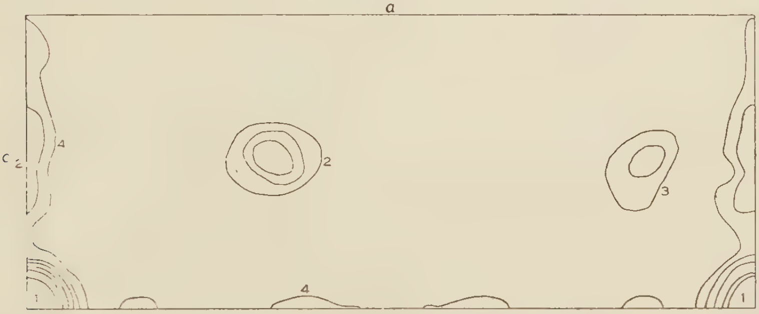


Fig. 1.—Fourier projection of K_2PtF_6 on (010) plane.
Peak 1, ¼ platinum atom. Peak 2, 1 potassium atom.
Peak 3, 2 fluorine atoms. Peak 4, diffraction ripples.

IV. Description of the Structure

The crystal is essentially a close packed arrangement of potassium and hexafluoroplatinate ions. Whether a compound of the type R_2MX_6 adopts the hexagonal (K_2GeF_6) (Hoard and Vincent 1939) or the cubic Cs_2GeF_6 structure depends on the relative sizes of R^{++} and $MX_6^{=}$. The fact that the $PtF_6^{=}$ complex is smaller than $PtCl_6^{=}$ causes K_2PtF_6 to adopt the hexagonal structure of K_2GeF_6 rather than the cubic structure of K_2PtCl_6.

Within a PtF_6^{-} complex, the F-F distances fall into two groups—2·66 and 2·63 Å respectively (see Fig. 2), so that the octahedron is slightly irregular. The F-F distance 2·66 Å agrees with the generally accepted van der Waals (and ionic radius) of F^{-} (1·33 Å).

The Pt-F bond has a length of 1·91 ± 0·02 Å which is appreciably smaller than the sum of the covalent radii of fluorine (0·66 Å) and octahedrally coordinated platinum (IV) (1·32 Å). This deviation is doubtless due to the fact that the Pt-F bond has considerable ionic character as would be expected from the high electronegativity of fluorine. If the ionic radius of fluorine is taken as 1·33 Å, the Pt_{IV} ion has a radius of 0·58 Å.

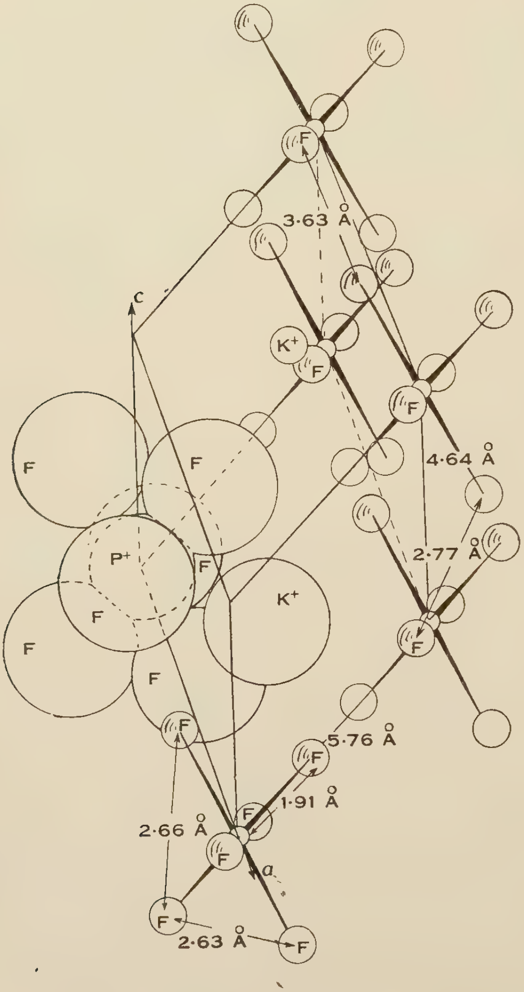


Fig. 2.—Diagram of K_2PtF_6 unit cell.

V. ACKNOWLEDGMENTS

The authors are indebted to Dr. A. McL. Mathieson for much helpful advice and to Dr. I. W. Wark for the use of facilities at the laboratories of the Division of Industrial Chemistry, C.S.I.R.O., during the latter part of the investigation.

VI. References

Emeleus, H. J., and Sharpe, A. G. (1950).—*J. Chem. Soc.* **1950** : 3444.

Hoard, J. L., and Vincent, W. B. (1939).—*J. Amer. Chem. Soc.* **61** : 2849.

Schlesinger, H. I., and Tapley, M. W. (1924).—*J. Amer. Chem. Soc.* **46** : 276.

Weisz, O., and Cole, N. (1948).—*J. Sci. Instrum.* **25** : 213.

THE CRYSTALLINE STRUCTURE OF CELLULOSE AS REVEALED BY ACID HYDROLYSIS

By D. H. Foster* and A. B. Wardrop*

[*Manuscript received April* 16, 1951]

Summary

The hydrolysis of holocellulose from *Eucalyptus regnans* F. v. M. has been studied employing a percolation technique and using 2 per cent. sulphuric acid for varying times of hydrolysis (0 to 6 hr.) and at three temperatures, 150, 160, and 170 °C. For comparison a series of experiments at 160 °C. on cotton was also carried out. The course of hydrolysis of the resistant cellulose in each experiment could be represented by the equation for a first order reaction. The " reaction constant " for cotton was smaller than that for holocellulose.

Powder X-ray diffraction photographs were made of samples at each stage of hydrolysis in the 160 °C. series. From these photographs the breadth at half maximum intensity of the line corresponding to the (002) planes, and the combined lines corresponding to the (101) and $(10\bar{1})$ planes was determined by photometric measurement. Throughout the period of hydrolysis studied, the line breadths decreased for both holocellulose and for cotton. This observation is considered to indicate the existence of a distribution of the degree of lateral order of the crystalline regions in each sample. The narrow line breadths in the cotton series may indicate the existence of greater lateral order in the cotton than in the wood holocellulose and may account for the observed smaller reaction constant for cotton. The results obtained are consistent with the view that the mesomorphous regions surrounding the micelles are more extensive in wood holocellulose than in cotton.

I. Introduction

Under conditions of high temperature (100 to 200 °C.) and pressure the dilute acid hydrolysis of cellulose can be effected either by autoclaving or by a percolation technique. In the latter method dilute acid is forced continuously or periodically through the cellulose under pressure at the desired temperature, the hydrolysis products being removed with the effluent acid (Luers 1930, 1932; Harris and Beglinger 1946.)

Saeman (1945) has demonstrated by means of the autoclaving technique that the course of wood cellulose hydrolysis can be described mathematically by a first order equation, although the application of this relationship was empirical. In a study of this kind it is clear that the hydrolysis reaction is complicated by the presence of degradation products of lignin and carbohydrates, and that the course of hydrolysis of the cellulose is likely to be influenced by the lignin present in the raw material. Because of these factors it is considered that the use of wood holocellulose as the raw material for hydrolysis by a percolation technique could be expected to give a clearer picture of the course of

---

* Division of Forest Products, C.S.I.R.O., Melbourne.

hydrolysis than is obtained by the autoclaving method. By using a percolation method it should be possible to remove hydrolysis products before they are able to form insoluble humic compounds.

It will also be apparent that in a heterogeneous reaction of this kind the structural changes in the cellulose will be of importance in determining the course of hydrolysis and the form of the hydrolysis curve. Accordingly in the present investigation X-ray examination of the hydrolysed cellulose has been made at each stage of hydrolysis. The percolation method of hydrolysis has also been adopted.

II. Experimental Procedure and Results

(*a*) *Materials Used*

The method of holocellulose determination was a quantitative adaptation of a method applied to thin wood sections (Kerr and Bailey 1934) and involves the use of alternate treatments with chlorine water and ammoniacal alcohol.*

The percentage composition of the original wood was:

Holocellulose	79·1†
Lignin	22·6
Pentosan	18·7
Uronic acid	5·0
Acetyl	3·7

The cotton used was obtained from a sample of carded cotton which had not been subjected to any previous chemical treatment. The cotton was extracted with ether and alcohol and air-dried. This was followed by boiling with 1 per cent. sodium hydroxide for 2 hours, washing in running water, followed by a washing in 1 per cent. acetic acid, and a final thorough water washing. The cotton was then air-dried.

(*b*) *Details of Percolator and Hydrolysis Technique*

The dilute sulphuric acid was forced by nitrogen gas pressure through a heated copper coil and then through the sample which was contained in the percolator tube shown in Figure 1. Both the coil and the tube were heated in a temperature controlled oil-bath.

To commence an experiment, the cellulosic material (2 g., air-dry) was placed in an alundum thimble which was set in place in the percolator tube and held there by a spring and threaded copper ring. The cap and valve were screwed on to the bottom of the tube, and after the thimble and sample had been warmed in the tube for 4 to 5 minutes, acid was allowed to run in very slowly. This was taken as the time at which hydrolysis commenced although a further 8 minutes was required to reach within 2 to 3 °C. of the desired temperature.

Acid was run away from the percolator tube at the rate of 25 ml. every 3 minutes. The nitrogen pressure was always held at 140 lb./sq. in. Sulphuric acid of 2 per cent. concentration was used in all experiments.

---

* Unpublished data (W. E. Cohen and T. M. Reynolds, Division of Forest Products).

† Containing 44·9 per cent. α-cellulose on original wood basis.

To terminate the hydrolysis the percolator was emptied of acid, the sample removed and rapidly cooled with water. This took about 5 minutes. The residue was then transferred to a tared alundum crucible, washed* with 400 ml. of distilled water, dried, and weighed.

With this technique three series of holocellulose samples were treated at 150, 160, and 170 °C. respectively for various time periods. A comparative series of experiments on cotton at 160 °C. was also carried out. The results of these experiments are summarized in Tables 1 and 2.

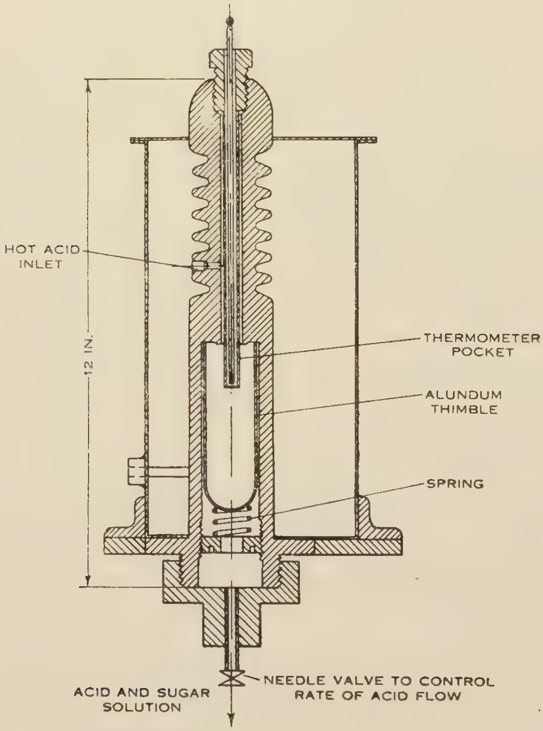


Fig. 1.—Cross section of portion of pressure-percolation apparatus.

There was no discoloration of the cellulose residues from the holocellulose with the exception that in the longer cooks a slight brown colour developed which was probably due to the concentration of a small amount of lignin present in the original holocellulose. Similarly the cotton was not discoloured during hydrolysis with the exception of a slight grey colour which was noticed in the 5- and 7-hour cooks.

---

* It has been discovered (Ranby 1950) that peptization of acid hydrolysed cellulose occurs at about pH 4. However, experiments carried out by the authors have shown that no serious loss of cellulose is involved during the water wash described here.

Table 1

HYDROLYSIS OF WOOD HOLOCELLULOSE WITH 2 PER CENT. SULPHURIC ACID

Residual cellulose as a percentage of original moisture-free holocellulose

150 °C.		160 °C.		170 °C.	
Time (hr.)	Residue (%)	Time (hr.)	Residue (%)	Time (hr.)	Residue (%)
2	44·9	1	41·8	0·5	38·8
4	35·8	2	31·4	1·0	24·3
6	30·1	4	14·1	1·3	17·6
8	21·9	6	7·5	2·0	9·4
10	18·2	8	3·2		
12	13·7	9·8	2·7		
14	12·0				
16	9·27				

Hence the residues were considered to be sufficiently pure to base calculations on their weights and no correction for contamination by humic substances was considered necessary.

Table 2

HYDROLYSIS OF COTTON CELLULOSE WITH 2 PER CENT. SULPHURIC ACID AT 160 °C.

Residual cellulose as a percentage of original moisture-free cellulose

Time (hr.)	Residue (%)
1	81·2
1·25	79·1
1·5	71·3
2	65·5
3	59·2
4	53·4
5	46·4
7	34·8

(c) *X-ray Examinations*

Powder X-ray diffraction photographs of the cotton and holocellulose samples after hydrolysis at 160 °C. were obtained. For this purpose CuK_α radiation was used with a cylindrical camera. The specimen-film distance was 7·16 cm. The breadths at half-maximum intensity of the 002 and combined

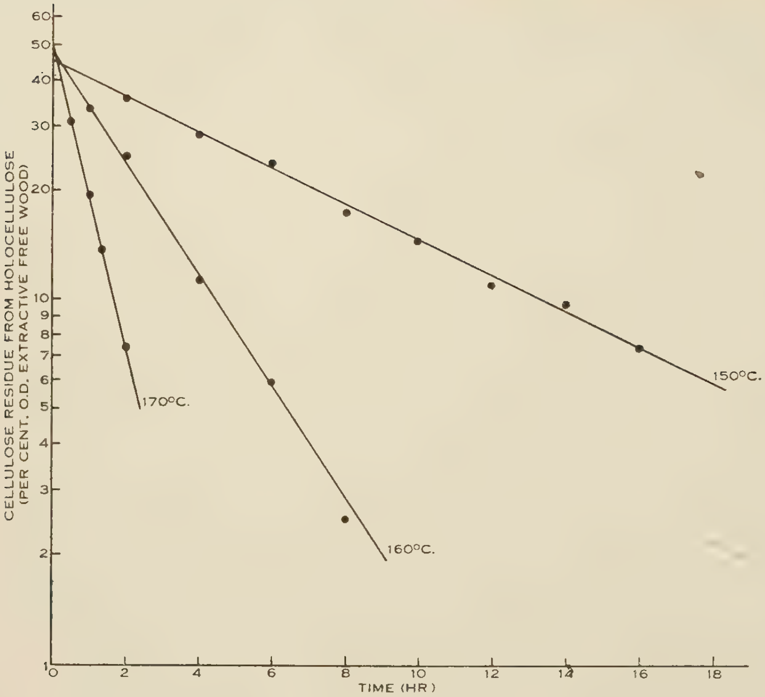


Fig. 2.—Holocellulose hydrolysis.

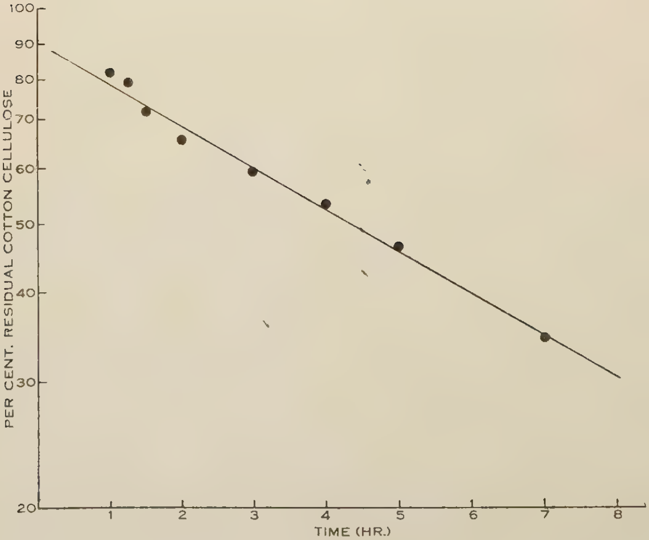


Fig. 3.—Cotton hydrolysis.

101 and $10\bar{1}$ diffraction lines were determined by photometric examination. After correcting for background intensity the values so obtained for holocellulose and cotton for different periods of hydrolysis are summarized in Table 3.

TABLE 3

BREADTH AT HALF-MAXIMUM INTENSITY (MM.) OF THE 002 AND COMBINED 101 AND $10\bar{1}$ DIFFRACTION LINES OF HYDROLYSED CELLULOSE SAMPLES

Specimen-film distance = 7·16 cm.

Time of Hydrolysis (min.)	Holocellulose (*E. regnans* F. v. M.)		Cotton	
	002	$101+10\bar{1}$	002	$101+10\bar{1}$
Control	3·75	4·80	2·30	4·35
18	3·30	4·65	—	—
60	2·90	4·30	1·70	4·25
90	2·80	—	—	—
120	2·80	4·20	1·90	4·25
180	2·70	4·45	—	—
240	—	—	1·75	4·15
300	—	—	1·80	4·10
360	2·50	4·10	—	—

III. DISCUSSION

In Figure 2 the residual holocellulose weights have been converted to a percentage of the original wood and plotted on a logarithmic scale against time of hydrolysis in hours. In Figure 3 similar results are plotted for the cotton hydrolysis at 160 °C. From the logarithmic plots it can be seen that straight lines are obtained as in a first order reaction. This result is similar to those obtained by other workers (Saeman 1945 ; Harris and Lang 1947 ; Harris and Kline 1949 ; Philipp, Nelson, and Ziifle 1947) who applied dilute acids to cellulosic materials without percolation.

From the slopes of the lines first order reaction constants were calculated and found to be :

	150 °C. (min.^{-1})	160 °C. (min.^{-1})	170 °C. (min.^{-1})
Holocellulose	0·00195	0·00572	0·0154
Cotton	—	0·0022	—

Thus for hydrolysis of the resistant cellulose fraction of the holocellulose a 10 °C. temperature rise between 150 and 170 °C. causes an average increase in the reaction constant by a factor of 2·80. This corresponds to an activation energy, calculated by the Arrhenius equation, of 38 kcal./mol. which may be compared with 35 kcal./mol. obtained by Philipp, Nelson, and Ziifle (1947) with hydrochloric acid at 95 and 100 °C. on cotton linters and on Fortisan rayon and with 43 kcal./mol. obtained by Saeman (1945) for the hydrolysis of wood cellulose with sulphuric acid at 170 to 190 °C. The differences may be due to the different temperatures at which activation energies were determined (Philipp, Nelson, and Ziifle 1947).

Harris and Lang (1947) consider that the measured activation energy includes that required to render the cellulose chains soluble as well as the activation energy for the hydrolytic fission of the 1–4β-glucosidic links.

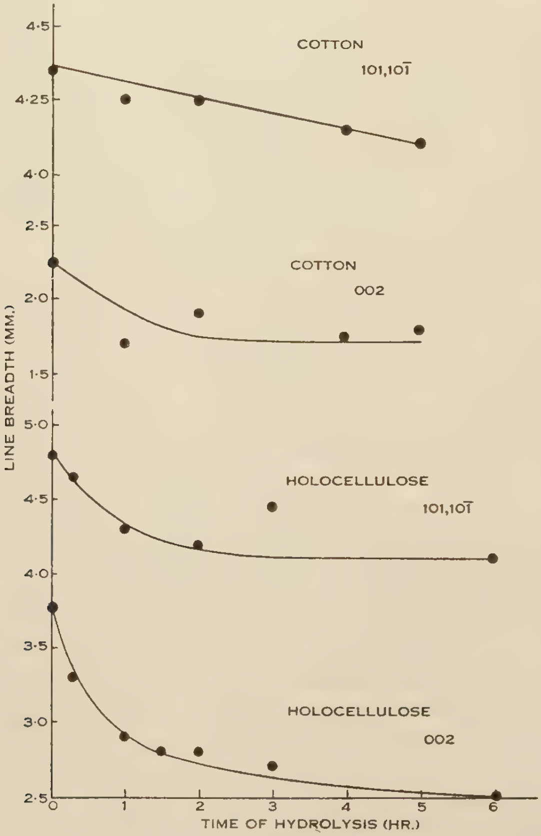


Fig. 4.—Change in breadth at half-maximum intensity of the 002 and combined 101 and 10$\bar{1}$ diffraction lines with time of hydrolysis for cotton and holocellulose of *Eucalyptus regnans*.

Extrapolation to zero time of the plots of the logarithm of residual cellulose against time of hydrolysis gives values ranging from 45·0 to 49·5 per cent. of the original wood (α-cellulose in holocellulose=44·9 per cent.). Although this is only an approximate figure, it may be considered to represent resistant or crystalline cellulose which may include some material crystallized during hydrolysis. Howsmon (1949) considered that hydrolysis with boiling 2·4N HCl : 0·6M ferric chloride caused crystallization of 9 per cent. of a wood pulp and 34 per cent. of a viscose rayon, and that most additional crystallization took place during the early stages of hydrolysis. Battista (1950) suggested that when 1–4 glucoside bonds are split at a slow rate (5N HCl, 18 °C.), longer and less acid-soluble crystallized material is obtained. More severe conditions (boiling 2·5N HCl) on the other hand promote the formation of very short and more acid-soluble crystalline material.

However, there is no evidence of the extent of crystallization in the range of 150 to 170 °C. and with dilute acid. Thus, although the more resistant part of the holocellulose appears to correspond in amount to the α-cellulose fraction, it must also contain any material crystallized during hydrolysis.

An examination of Figure 4 shows that the 002 and combined 101 and $10\bar{1}$ diffraction lines become sharper in progressive stages of hydrolysis both for cotton and for the holocellulose samples, although the effect is more exaggerated in the latter case. In a recent paper Hermans and Weidinger (1949) report that for ramie only the combined 101 and $10\bar{1}$ diffraction line increases in sharpness during hydrolysis, and the 002 line remains of constant breadth. This conclusion was reached apparently upon the examination of only three samples, and is not consistent with the evidence of Figure 4.

Accepting the constancy of the 002 line, however, Hermans and Weidinger conclude that the sharpening of the 101 and $10\bar{1}$ lines is to be attributed to the formation of cellulose IV, in the X-ray diffraction diagram of which the lines 101 and $10\bar{1}$ merge into a single line.

However, in view of the results obtained in the present investigation it appears improbable that the explanation offered by Hermans and Weidinger is correct. Thus the combined 101 and $10\bar{1}$ and the 002 lines showed considerable sharpening both for cotton and for holocellulose and the 101 and $10\bar{1}$ lines were separately distinguishable in the most severely hydrolysed samples. Also it would require the conversion to cellulose IV of a considerable fraction of the cellulose present for this to be apparent in the X-ray diagram. In view of these observations an alternative explanation of the data is desirable.

If the sharpening of the X-ray diffraction lines is taken to indicate an increase in the degree of lateral order of the micelles, then any crystallization which occurs must contribute to this effect. However, the molecules potentially capable of crystallization during hydrolysis are also the ones most likely to be hydrolysed, so that any increase in the extent of the crystalline regions due to crystallization might be expected to occur in the initial stages of hydrolysis as has been suggested by a number of investigators (Howsmon 1949 ; Wardrop 1949). Thus although crystallization may account for part of the sharpening of the diffraction lines observed during the initial stages of hydrolysis it is unlikely that such a conception offers an adequate explanation of the increasing sharpness of the diffraction lines throughout the period of hydrolysis studied above.

An alternative explanation to the foregoing is that there is a range of size of the micelles, and during hydrolysis the regions of lowest lateral order are removed first so that the average degree of lateral order of the micelles did not decrease. Such a conception is implicit in the observation that X-ray methods of estimating the crystalline/non-crystalline ratio of cellulose give lower values than do chemical methods. This arises from the fact that for particles of less than about 25 Å, the X-ray diffraction pattern resembles more and more that of a liquid and these regions of extremely limited lateral order are estimated as part of the non-crystalline fraction of the cellulose. Relative to the assumption of the existence of a micelle size distribution, it has been calculated by E. J. Williams (personal communication) that if cellulose crystallites are attacked only on the lateral faces, then in order to simulate a first order reaction there must be an exponential distribution of micelle breadths. Hence there should be a large number of micelles of small breadth and relatively few of larger lateral dimensions. This is of interest in relation to a similar conclusion by Hosemann (1940) on the basis of a study of small angle X-ray scattering, although these results have been subsequently criticized by Kratky, Sekora, and Treer (1942).

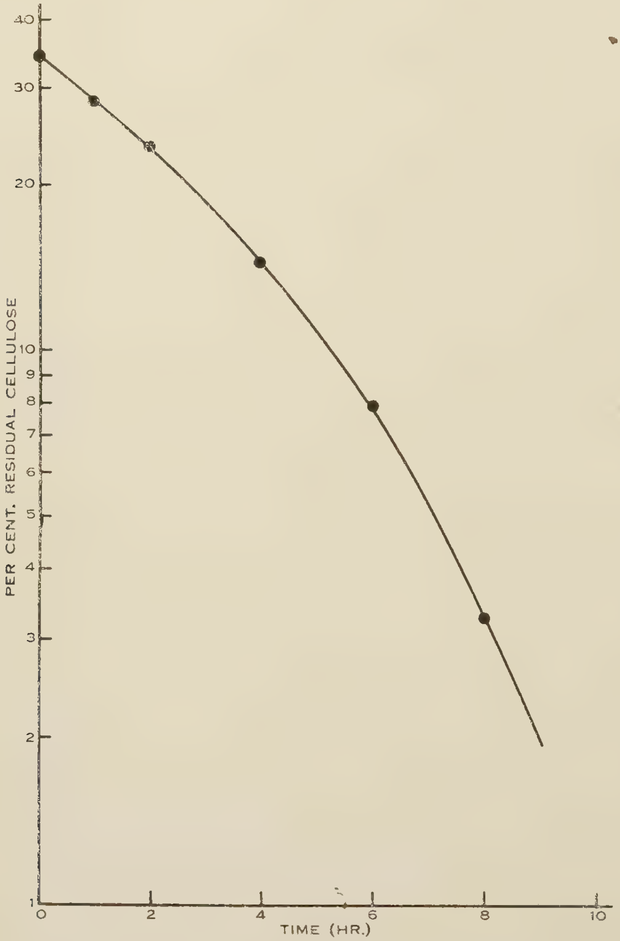


Fig. 5.—Calculated form of hydrolysis curve if all cellulose crystallites were of the same size and shape.

It may be noted further that if the micelles were of uniform breadth then the progressive change in their size during hydrolysis because of the increase in the ratio of surface to volume would lead to an increasing reaction rate with time as illustrated in Figure 5 (E. J. Williams, personal communication). This may be compared with the actual hydrolysis in Figure 2. A difference in average micelle breadth may also explain the difference in hydrolysis rates between cotton and wood holocellulose (Figs. 2 and 3 ; see also Howsmon 1949). The line breadth in cotton indicates that the micelles are here relatively broader than in the holocellulose, so that upon the basis of the above discussion a slower hydrolysis rate might be anticipated.

During the initial stages of hydrolysis the greater sharpening of the diffraction line observed for the holocellulose sample as compared with the cotton (Fig. 4) suggests that more crystallization occurs in the former sample. This is consistent with the work of Jorgensen (1949) from which it was concluded that the extent of the mesomorphous regions was greater in wood fibres than in cotton.

In the initial stages of hydrolysis short-chain non-oriented molecules are removed from between the micelles. The removal of this amorphous material is responsible for initial rapid hydrolysis rate and in part is reflected by a sharpening of the X-ray diffraction lines. This stage is followed or accompanied by an attack on the oriented non-crystalline or mesomorphous longer-chain molecules which gradually merge into the crystalline material of the micelles. During this stage some of the long-chain molecules pass into solution, while others may rearrange and crystallize and so lead to an increase in the lateral and longitudinal order of the crystalline regions. Finally the long-chain molecules arranged in the crystalline lattice of the micelles are hydrolysed.

IV. Acknowledgments

The authors wish to thank Mr. E. J. Williams, Section of Mathematical Statistics, C.S.I.R.O., for his assistance in solving the mathematical problems associated with the work. Acknowledgment is also due to Mr. L. Santer and Dr. B. J. Ralph for their part in designing the percolator.

V. References

Battista, A. (1950).—*Industr. Engng. Chem.* **42** : 502-7.

Harris, E. E., and Beglinger, E. (1946).—*Industr. Engng. Chem.* **38** : 890-5.

Harris, E. E., and Kline, A. A. (1949).—*J. Phys. Coll. Chem.* **53** : 344-57.

Harris, E. E., and Lang, B. G. (1947).—*J. Phys. Coll. Chem.* **51** : 1430-41.

Hermans, P. H., and Weidinger, A. (1949).—*J. Polymer Sci.* **4** : 317.

Hosemann, R. (1940).—*Z. Angew. Chem.* **53** : 332. Quoted by Hermans, P. H. (1949).—" The Physics and Chemistry of Cellulose Fibres." (Elsevier Publishing Co. Inc. : New York.)

Howsmon, J. A. (1949).—*Text. Res. J.* **19** : 153-62.

Jorgenson, L. (1949).—*Acta Chem. Scand.* **3** : 780-2.

Kerr, T., and Bailey, I. W. (1934).—*J. Arnold Arbor.* **15** : 327-49.

Kratky, O., Sekora, A., and Treer, R. (1942).—*Z. Electrochem.* **48** : 587.

Luers, H. (1930).—*Z. Angew. Chem.* **43** : 458.

Luers, H. (1932).—*Z. Angew. Chem.* **45** : 376.

Philipp, H. J., Nelson, M. L., and Ziifle, H. M. (1947).—*Text. Res. J.* **17** : 585-96.

Saeman, J. F. (1945).—*Industr. Engng. Chem.* **37** : 43-52.

Rånby, B. J. (1950).—*Acta Chem. Scand.* **3** : 649-50.

Wardrop, A. B. (1949).—*Nature* **164** : 366.

EXPERIMENTS ON THE SYNTHESIS OF THE ACRIDONE ALKALOIDS

By G. K. Hughes* and E. Ritchie*

[*Manuscript received March* 6, 1951]

Summary

A scheme for the biosynthesis of the acridone alkaloids is suggested and experimental evidence bearing on this question is described.

I. Introduction

As a result of the work of Price (1949, personal communication), Hughes (1949, 1950, unpublished work), Lahey (1949), and their respective collaborators on the acridone alkaloids, two generalizations can be drawn. Firstly the substituents are present in only one of the two benzene nuclei and secondly the 2- and the 4-positions are always substituted. Thus the simplest member of the group is 2,4-dimethoxy-10-methylacridone, which has been isolated from the leaves of *Acronychia baueri* Schott.

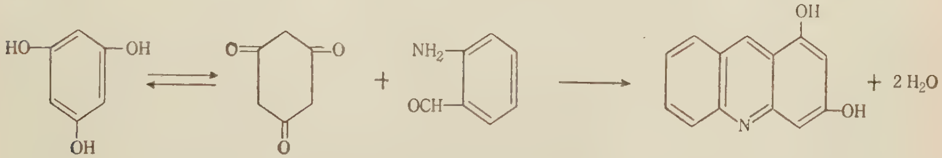
On the basis of these facts a scheme for the biosynthesis of the acridone alkaloids is now suggested. The reactions postulated are:

(i) The condensation of *o*-aminobenzaldehyde (or its equivalent) with phloroglucinol (or its equivalent) to form 2,4-dihydroxyacridine (I).

(ii) Oxidation of the acridine to an acridone.

(iii) Nuclear oxidation to 2,3,4-trihydroxy- and 1,2,3,4-tetrahydroxyacridone.

(iv) Methylation (and sometimes methylenation) to give the *O* and *O,N* derivatives.



(I)

It is not thought that the reactions necessarily occur in the above order and it is obvious that several alternatives are equally likely. For example, it is possible that a hydroxy derivative of phloroglucinol may be involved in the reactions leading to evoxanthine or melicopicine, and that methylation of the polyhydroxy acridine to a quaternary salt followed by oxidation, may be the route by which the nitrogen is methylated and acridone formation achieved.

Experimental confirmation of this scheme was sought. The reaction between phloroglucinol and *o*-aminobenzaldehyde to form I has been described by Eliasberg and Friedlander (1892), who effected the condensation in aqueous solution on the water-bath by the addition of a few drops of sodium hydroxide solution. They described the product as a voluminous red substance, soluble in both acid and alkali but sparingly soluble in almost all solvents and very difficult to purify by recrystallization. No melting point was reported but the substance was characterized by its crystalline chloroplatinate and dibenzoate (m.p. 163 °C.). Repetition of these experiments confirmed their results but it was also found that the condensation proceeded very readily under quite mild conditions. The yield of 2,4-dihydroxyacridine obtained from equimolecular proportions of the reactants in dilute aqueous solution at room temperature for five days varied with the pH as follows:

---

* School of Chemistry, University of Sydney.

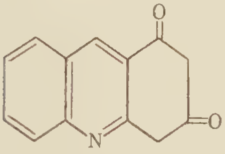
pH ..	4	5	6	7	8	9	10	11	12	13
Yield (%) ..	0	0	5	28	90	87	87	83	81	43

Thus the reaction proceeds readily under conditions much milder than those found necessary by Schöpf and Lehmann (1932) in the analogous reaction between *o*-aminobenzaldehyde and other ketones. From the reaction at pH 8 the pure product crystallized out in very deep reddish brown needles which analysed for a sesquihydrate, but all efforts to recrystallize this or any other sample were circumvented by its remarkably low solubility, its slight temperature gradient, and its tendency to separate in the amorphous state.

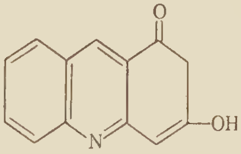
As attempts to convert I directly to a trihydroxyacridine by an Elb's persulphate oxidation (Baker and Brown 1948) were unsuccessful, alternative paths were explored. Because of its low solubility it was impracticable to attempt to form a quaternary salt from I, so the use of more soluble derivatives was examined. The dibenzoate of I (m.p. 170 °C.), obtained in rather low yield by the action of a large excess of benzoyl chloride on a cold alkaline solution of I, was converted to its *methosulphate* which was then oxidized with alkaline ferricyanide. Instead of the expected 2,4-dihydroxy-10-methylacridone, a deep reddish brown *substance*, $C_{15}H_{13}O_3N$, not further examined, was formed in low yield together with much amorphous material.

Secondly the methylation of I was investigated. With alkali and dimethyl sulphate, even in very large excess, I gave only low yields of *2,4-dimethoxyacridine* and none of the starting material could be recovered. This low yield and the relative difficulty of benzoylation are undoubtedly due to tautomerism of I involving the forms I, II, III, IV, V, and VI. Spectroscopic evidence for the existence of tautomeric forms analogous to VI and V in 2- and 4-hydroxyacridines respectively has been adduced by Albert and Short (1945), and evidence for the existence of II, III, and IV is provided by the condensation of phloroglucinol with two and three molecular proportions of *o*-aminobenzaldehyde to form polynuclear derivatives of acridine (Niementowski 1906; Borsche, Wagner-Roemmich, and Barthenheier 1942). Methylation with diazomethane, however, gave a nearly quantitative yield of 2,4-dimethoxyacridine, the *methosulphate* of which was oxidized by hot alkaline ferricyanide in nearly quantitative yield to 2,4-dimethoxy-10-methylacridone, identified by comparison with an authentic specimen. The result of this oxidation experiment is contrary to the findings of Lehmstedt and Hundertmark (1931) who, after a study of the oxidation of acridine methosulphate with this reagent, concluded that the ferricyanide took no part in the reaction and that the 50 per cent. yield of 10-methylacridone they obtained was formed by simple disproportionation of the carbinol base.

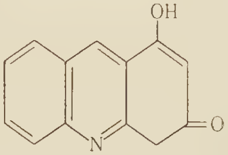
Experiments were then carried out to convert the " nor " compound, 2-methoxy-4-hydroxy-10-methylacridone, to a trihydroxy derivative, but these were uniformly unsuccessful. The action of hexamine in glacial acetic acid (Rao, Seshadri, and Thiruvengadam 1949) gave a complex mixture of basic products, not further examined. Oxidation with lead tetra-acetate (Fieser and Hershberg 1938) at 100 °C. or at room temperature gave much amorphous material plus a little unchanged starting material, and a similar result was obtained with hydrogen peroxide and ferrous sulphate (Merz and Waters 1949). With alkaline potassium persulphate some starting material could be recovered and a small yield of a reddish brown *substance*, m.p. 235 °C. (decomp.) of unknown constitution, was isolated. By the action of hydrobromic acid the " nor " compound gave 2,4-dihydroxy-10-methylacridone, but attempts to oxidize this with cold nitric acid or with selenium dioxide in boiling ethyl acetate to a hydroxyquinone or with alkaline persulphate were again unsuccessful. The latter reagent gave much amorphous material and a small yield of a reddish brown *substance* of unknown structure.



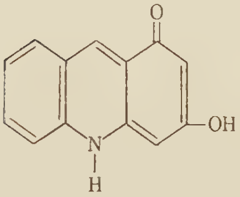
(II)



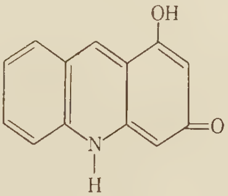
(III)



(IV)

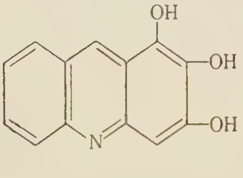


(V)

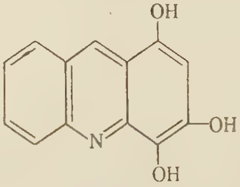


(VI)

Two other routes to the desired trihydroxy compound were examined. From the reaction of phloroglucinaldehyde with *o*-aminobenzaldehyde it was hoped to obtain 2,4-dihydroxy-3-formylacridine which after methylation should be oxidizable to a derivative of 2,3,4-trihydroxyacridine. The product, a red amorphous uncrystallizable powder, did not analyse correctly, nor could it be methylated, so this line was abandoned. The condensation between *o*-aminobenzaldehyde and 1,2,3,5-tetrahydroxybenzene (Posternak 1936) could theoretically give VII and/or VIII, but as shown below only VII was formed. At pH 8 at room temperature for 5 days much amorphous material was formed and the yield of VII was minute but, in hot aqueous solution in the presence of one molecular proportion of potassium hydroxide, the reaction was much cleaner, in spite of the extreme sensitivity of 1,2,3,5-tetrahydroxybenzene to alkali. The crude products, which were uncrystallizable, were immediately methylated with diazomethane. The product was a single substance *2,3,4-trimethoxyacridine.*



(VII)



(VIII)

It readily formed a *methosulphate* which was oxidized by alkaline ferricyanide in nearly quantitative yield to 2,3,4-trimethoxy-10-methylacridone, identified by comparison with an authentic specimen (Hughes 1950).

Finally it may be pointed out that the ability of phenols to condense with *o*-aminobenzaldehyde to an acridine derivative appears to be limited to those with three hydroxyl groups in the 1-, 3-, and 5-positions, since phenol, salicylic acid, catechol, resorcinol, β-resorcylic acid, quinol, α-naphthol, β-naphthol, and pyrogallol all failed to react.

II. EXPERIMENTAL

(*a*) *2,4-Dihydroxyacridine* (*I*)

To a solution of *o*-aminobenzaldehyde (0·6 g.; 1 mol.) in a little alcohol was added anhydrous phloroglucinol (0·63 g.; 1 mol.) and the appropriate buffer solution (250 cc.) which had been just previously boiled and cooled in an atmosphere of nitrogen. The flask, which was then completely filled, was immediately fitted with an air-tight stopper and shaken thoroughly. In the experiments at pH's 13 to 10 inclusive the reaction mixture rapidly became deep red and this colour appeared to reach its maximum intensity in a few hours. At pH 9 the colour developed more slowly but in 24 hours the product began to separate. At pH 8 the red colour was slower still in reaching its maximum intensity but again the product began to crystallize out after 24 hours. At lower pH's the colour did not become deeper than reddish orange although amorphous precipitates formed in a few hours.

After being allowed to stand at room temperature for 5 days, the pH was adjusted to 6–8, the product collected, washed with water (and with warm acetone in the experiments at pH 4–7 to remove polymers of *o*-aminobenzaldehyde), and dried at 110 °C. The yields have been given above.

A preparation, which had crystallized in small deep red needles from a reaction at pH 8, was collected, washed with water, dried overnight in an evacuated desiccator over concentrated sulphuric acid, and analysed.

Found: C, 65·5; H, 4·9; N, 5·9%.

Calculated for $C_{13}H_9O_2N,1\frac{1}{2}H_2O$: C, 65·7; H, 5·1; N, 5·9%.

After drying in a vacuum pistol at 100 °C./1 mm. for 6 hours, the substance had lost weight but it took up moisture so rapidly during weighing that an accurate determination of water of crystallization was impossible. Similar behaviour was observed in a sample prepared by the hydrolysis of the pure dibenzoyl derivative (see below).

The substance dissolved readily in dilute mineral acids and in dilute sodium carbonate to give deep red solutions. It was very sparingly soluble in most organic solvents including boiling pyridine and nitrobenzene and in water but such solutions were deep red in spite of their low concentrations. On heating it began to char at about 315 °C. but did not melt.

For ordinary purposes the substance was prepared as follows. To a solution of *o*-aminobenzaldehyde (7·2 g. ; 1 mol.) in a little alcohol was added anhydrous phloroglucinol (7·5 g. ; 1 mol.) and sodium hydroxide (2·4 g. ; 1 mol.) dissolved in water (500 cc.). The deep red solution was boiled for 5 minutes and then glacial acetic acid (5 cc.) diluted with water run into the hot solution with vigorous stirring. The bright red voluminous precipitate was collected, washed with water, and dried at 110 °C. (Yield 90%.)

(*b*) *2,4-Dibenzoyloxyacridine*

Since the description given by Eliasberg and Friedlander (1892) is incomplete, the preparation is described fully.

Benzoyl chloride (100 g.) was added in 5 portions with vigorous shaking and cooling in ice to a solution of I (12 g.) and sodium hydroxide (54 g.) in ice water (500 cc.) during 30 minutes. Although the colour of the reaction mixture did not change appreciably, a brown precipitate gradually formed. When the odour of benzoyl chloride had disappeared, the product was collected (starting material could not be recovered from the alkaline filtrate), triturated with dilute sodium hydroxide, washed thoroughly with water, dried, and extracted with boiling benzene (2 × 200 cc.). The filtered extract was concentrated to about 100 cc., cooled, and light petroleum (b.p. 60–80 °C.) added to incipient turbidity. After adding charcoal and filtering, the liquid was evaporated to dryness. The residue was dissolved in hot alcohol (300 cc.), concentrated hydrochloric acid (10 cc.) added, the liquid cooled, the bright yellow crystalline hydrochloride collected and washed with a little cold alcohol. The base, regenerated by shaking with dilute alkali and benzene, was eventually isolated after several recrystallizations from acetone as very pale yellow hard prisms, m.p. 170 °C. (10 g. ; 40%).

Found : C, 77·4 ; H, 4·1%.

Calculated for $C_{27}H_{17}O_4N$: C, 77·3 ; H, 4·1%.

The substance is polymorphic, separating from ethyl acetate or benzene in needles and occasionally from acetone also, but on standing, the needles change to prisms. When pure it was fairly readily soluble in hot acetone or benzene but only slightly soluble in hot alcohol or ethyl acetate. Slightly impure samples dissolved quite readily in all of these solvents. Its solution in glacial acetic acid had a strong greenish yellow fluorescence.

Hydrolysis was effected by refluxing it (4 g.) with 10% aqueous alcoholic sodium hydroxide (100 cc.) for 3 hours. After evaporating the alcohol, the residue was dissolved in water (300 cc.), the solution filtered, heated to boiling and slowly acidified with glacial acetic acid (17 cc.) with vigorous stirring. The bright red product was filtered from the hot solution and washed with boiling water. It was purified by reprecipitation from hot alkaline solution, forming a bright red powder (1·9 g. ; 95%). On drying in a pistol this product showed the same behaviour as the sample prepared at pH 8.

(*c*) *2,4-Dibenzoyloxy-10-methylacridinium Methyl Sulphate*

A solution of the dibenzoate (9 g.) in pure dry benzene (200 cc.) and dimethyl sulphate (20 cc.) was heated under reflux in a dry atmosphere. In a short time the solution became bright yellow and the product began to separate. After 5 hours the mixture was cooled, the product collected and washed thoroughly with benzene. It formed bright yellow needles (9·9 g. ; 85%) which began to darken at 170 °C. and then decomposed at 180 °C. It was soluble in hot water but on cooling, the solution set to a jelly. In hot methanol or ethanol decomposition occurred, since the odour of benzoic esters was quite strong. It was insoluble in other solvents.

Found : C, 64·1 ; H, 4·2%.

Calculated for $C_{29}H_{23}O_8NS$: C, 63·9 ; H, 4·2%.

(*d*) *Oxidation of the Quaternary Salt*

A suspension of the salt ($9 \cdot 7$ g.) in a solution of potassium ferricyanide ($14 \cdot 2$ g.; 20% excess) in water (150 cc.) was stirred and treated during 45 minutes with sodium hydroxide ($6 \cdot 8$ g.) in water (50 cc.). The mixture soon became deep red and the salt dissolved. After 3 hours the mixture was filtered and just acidified with acetic acid. The black precipitate was collected, washed, and extracted with boiling water (5 × 750 cc.). The crystalline product, which separated after concentrating to 200 cc. and cooling, was boiled with a little concentrated hydrochloric acid. The bright yellow hydrochloride, which was collected after cooling, was reconverted to the base by treatment with warm aqueous sodium acetate. It was finally purified by dissolving it in a large volume of hot water and concentrating to a small bulk. It formed deep reddish brown needles ($0 \cdot 4$ g.) which began to decompose at about 250 °C. It was soluble in dilute acids and alkalis but sparingly soluble in the usual organic solvents.

Found*: C, $71 \cdot 2$; H, $5 \cdot 1$; N, $5 \cdot 5$%.

Calculated for $C_{15}H_{13}O_3N$: C, $70 \cdot 6$; H, $5 \cdot 1$; N, $5 \cdot 5$%.

(*e*) *Methylation of I*

(i) Dimethyl sulphate ($37 \cdot 8$ g.) was added in 5 portions during 30 minutes with vigorous shaking and cooling in ice water to a cold solution of dihydroxyacridine ($6 \cdot 3$ g.) and sodium hydroxide ($12 \cdot 5$ g.) in water (200 cc.). Then more sodium hydroxide (12 g.) in water (50 cc.) was added, followed by dimethyl sulphate (34 g.) again added during 30 minutes. The mixture was warmed to 50 °C. for 10 minutes, cooled, and extracted with benzene. After washing with dilute alkali, the benzene was evaporated and the residue extracted with hydrochloric acid (2 × 80 cc. of N). The acid solution was basified and shaken with benzene (200 cc.). The extract, after drying, was passed through a short column of alumina when the dark impurities were preferentially adsorbed. Evaporation of the benzene eluates gave a nearly pure product ($1 \cdot 4$ g.; 20%). If cooling was omitted during the addition of the dimethyl sulphate, the yield was slightly lower and the product more difficult to purify. No pure substance could be isolated from the alkaline liquors.

(ii) A suspension of finely powdered dihydroxyacridine (6 g.) in methanol (50 cc.) was stirred and cooled in ice. Ethereal diazomethane (400 cc. from 40 g. of nitrosomethylurea) was added and stirring continued until all the solid had disappeared (3–4 hours). After standing overnight the red solution was evaporated to dryness, the residue extracted with hot dilute acid, the red extract filtered, and basified. The crude product ($6 \cdot 9$ g.), obtained by benzene extraction, was sufficiently pure for the next step. For analysis it was purified by passing its benzene solution through alumina followed by recrystallization from light petroleum (b.p. 60–80 °C.) containing a little acetone. It formed small yellow needles, m.p. 114 °C.

Found: N, $6 \cdot 0$; CH_3O, $26 \cdot 0$%.

Calculated for $C_{15}H_{13}O_2N$: N, $5 \cdot 9$; $2 \times CH_3O$, $25 \cdot 9$%.

It was sparingly soluble in light petroleum but readily soluble in other organic solvents giving pale yellow solutions with a strong blue fluorescence. It dissolved in dilute mineral acids and in glacial acetic acid to red non-fluorescent solutions.

(*f*) *2,4-Dimethoxy-10-methylacridinium Methyl Sulphate*

When dimethyl sulphate (5 cc.) was added to a warm solution of crude dimethoxyacridine (3 g.) in dry benzene (25 cc.), a crystalline product began to separate almost immediately. After warming on the water-bath for 30 minutes, the mixture was cooled, the product collected, and washed with benzene and acetone. It crystallized from alcohol in almost quantitative yield in orange-yellow needles, which began to decompose at about 235 °C. and effervesced at 242 °C.

Found: C, $55 \cdot 7$; H, $5 \cdot 2$%.

Calculated for $C_{17}H_{19}O_6NS$: C, $55 \cdot 9$; H, $5 \cdot 2$%.

(*g*) *2,4-Dimethoxy-10-methylacridine*

A solution of potassium ferricyanide (8 g.; 20% excess) in water (50 cc.) was stirred and heated on the water-bath while a solution of the above quaternary salt ($3 \cdot 6$ g.) in hot water

(30 cc.) was run in. The clear deep red solution was treated during 30 minutes with a solution of potassium hydroxide (3 g.) in water (25 cc.), when a solid product gradually separated. After 2 hours the mixture was cooled and extracted with chloroform. The extract was washed successively with water, 1% acetic acid (to remove any traces of carbinol base), water, dilute alkali, and water, then dried and evaporated. The residue (yield, nearly quantitative), crystallized from benzene-light petroleum (b.p. 60–80 °C.) in very pale yellow needles, m.p. 163–165 °C., undepressed by admixture with the natural product.

Found: C, 71·4; H, 5·7%.

Calculated for $C_{16}H_{15}O_3N$: C, 71·3; H, 5·6%.

When the oxidation was effected at room temperature for 8 hours the yield dropped to 25%.

The picrate crystallized from methanol in yellow needles, m.p. 203–205 °C., undepressed by admixture with an authentic specimen.

Found: N, 11·3%.

Calculated for $C_{22}H_{18}O_{10}N_4$: N, 11·2%.

The "nor" derivative, prepared by keeping the hydrochloride of the base molten at 170–180 °C. for 5 minutes, crystallized from alcohol-chloroform in yellow prismatic needles, m.p. 175–176 °C., undepressed by admixture with an authentic specimen.

Found: C, 70·8; H, 5·2%.

Calculated for $C_{15}H_{13}O_3N$: C, 70·6; H, 5·1%.

(*h*) *Oxidation of the "Nor" Derivative*

A mixture of the "nor" derivative (1 g.; 1 mol.) dissolved in pyridine (40 cc.) and potassium hydroxide (1·1 g.) dissolved in water (10 cc.) was stirred and treated during 3 hours with a solution of potassium persulphate (1·1 g.; 1 mol.) in water (25 cc.), the temperature being kept below 20 °C. After standing overnight, concentrated hydrochloric acid (60 cc.) was added below 20 °C., the mixture warmed to 50 °C., and filtered from starting material (0·6 g.). Sodium sulphite (3 g.) and concentrated hydrochloric acid (30 cc.) were added, the mixture heated on the water-bath for 30 minutes, and then rapidly concentrated to half its bulk. On standing for several days, a dark crystalline precipitate (0·1 g.) appeared which, after crystallization in turn from alcohol, benzene, ethyl acetate and alcohol, formed small reddish brown needles which began to darken at 200 °C. and melted at 235 °C.

Found*: C, 66·3; H, 4·9; N, 6·4%.

(*i*) *2,4-Dihydroxy-10-methylacridone*

To a hot solution of the "nor" compound (2 g.) in glacial acetic acid (10 cc.) was added concentrated hydrobromic acid (30 cc.). The mixture, which set to a bright yellow paste, was refluxed for 90 minutes, during which time the undissolved material became greenish yellow. After pouring into water (600 cc.), the mixture was boiled to coagulate the precipitate, cooled, and filtered. The product was extracted with aqueous potassium hydroxide (200 cc. of 2%), the extract filtered, and saturated with carbon dioxide. The precipitate crystallized from alcohol in yellow prismatic needles (1·6 g.), which sintered at 275 °C. and decomposed at 288–292 °C. (lit. 286–289 °C.).

(*j*) *Oxidation of 2,4-Dihydroxy-10-methylacridone*

A solution of the substance (2·2 g.; 1 mol.) and potassium hydroxide (2·6 g.) in water (30 cc.) was treated with stirring with potassium persulphate (2·5 g.; 1 mol.) in water (70 cc.) during 3 hours, the temperature being kept below 20 °C. After standing overnight, the dark solution was acidified with concentrated hydrochloric acid (6 cc.) and filtered from the dark amorphous precipitate. Sodium sulphite (3 g.) and concentrated hydrochloric acid (50 cc.) were added to the combined filtrate and washings which were then heated on the water-bath for 30 minutes. The deep red crystalline precipitate, which had gradually formed, was collected after cooling and recrystallized several times from aqueous alcohol, forming reddish brown flattened needles (0·1 g.) which gradually charred on heating but did not melt below 300 °C.

Found*: C, 65·4; H, 4·9; N, 5·8%.

Attempts to methylate the substance with diazomethane gave only tarry products.

(*k*) *2,3,4-Trimethoxyacridine*

(i) *o*-Aminobenzaldehyde (0·42 g. ; 1 mol.) was condensed with 1,2,3,5-tetrahydroxybenzene (0·47 g. ; 1 mol.) in 250 cc. of pH 8 buffer at room temperature for 5 days as above. From the solution, which had immediately assumed a very dark brown colour, a dark amorphous precipitate slowly separated. On methylating with diazomethane, a yield of about 0·005 g. was obtained.

(ii) To a solution of *o*-aminobenzaldehyde (4·9 g. ; 1 mol.) in a little alcohol was added 1,2,3,5-tetrahydroxybenzene (5·5 g. ; 1 mol.) and air-free distilled water (140 cc.). The solution was quickly heated to the boiling point in an atmosphere of nitrogen and a solution of potassium hydroxide (2·2 g. ; 1 mol.) in water (10 cc.) added. After refluxing for 3 minutes, glacial acetic acid (3 cc.) followed by excess solid sodium acetate was added and the mixture cooled. The dark precipitate (7·5 g.) was collected, washed, and dried.

The crude product was methylated with diazomethane (from 70 g. of nitrosomethylurea) and the reaction mixture worked up as usual. After passing the benzene solution through alumina, yellow material (4·5 g. ; 40%), sufficiently pure for the next step, was obtained. The pure substance crystallized from light petroleum (b.p. <40 °C.) in long yellow needles, m.p. 62 °C. It was easily soluble in the usual organic solvents with a bluish green fluorescence.

Found : N, 5·2%.

Calculated for $C_{16}H_{15}O_3N$: N, 5·2%.

(*l*) *2,3,4-Trimethoxy-10-methylacridinium Methyl Sulphate*

The quaternary salt, prepared in nearly quantitative yield, by refluxing the crude base (5 g.) and dimethyl sulphate (10 cc.) in benzene (30 cc.) for 1 hour, crystallized from alcohol in orange needles, m.p. 220 °C. It is thermochromic, changing to deep red at temperatures below 100 °C. The change is reversible up to a few degrees below the melting point.

Found : C, 54·5 ; H, 5·5%.

Calculated for $C_{18}H_{21}O_7NS$: C, 54·7 ; H, 5·4%.

(*m*) *2,3,4-Trimethoxy-10-methylacridone*

Oxidation of the above quaternary salt (6 g.) by potassium ferricyanide (12·6 g.) and potassium hydroxide (5·2 g.) was effected on the water-bath as above. The product, obtained in nearly quantitative yield, crystallized from benzene-light petroleum (b.p. 60–80 °C.) in very pale yellow rhombs, m.p. 168–170 °C., undepressed by admixture with an authentic specimen.

Found : C, 68·2 ; H, 5·7%.

Calculated for $C_{17}H_{17}O_4N$: C, 68·3 ; H, 5·7%.

The " nor " derivative, obtained by refluxing the above substance (1 g.) in alcohol (60 cc.) and concentrated hydrochloric acid (3 cc.) for 2 hours, crystallized from alcohol in long fine golden needles, m.p. 176–177 °C., undepressed by admixture with an authentic specimen.

Found : C, 67·4 ; H, 5·2%.

Calculated for $C_{16}H_{15}O_4N$: C, 67·4 ; H, 5·3%.

III. ACKNOWLEDGMENTS

Analyses marked with an asterisk are due to Dr. Weiler and Dr. Strauss, Oxford ; the remainder were done by Mrs. E. Bielski, University of Sydney. The authors are grateful to Dr. S. J. Angyal for a supply of scyllo-inosose (for the preparation of 1,2,3,5-tetrahydroxybenzene) and to Dr. J. R. Price, C.S.I.R.O., for advance information on the alkaloids of the leaves of *A. baueri* and for a specimen of 2,4-dimethoxy-10-methylacridone.

IV. REFERENCES

ALBERT, A., and SHORT, L. N. (1945).—*J. Chem. Soc.* **1945** : 760.

BAKER, W., and BROWN, N. C. (1948).—*J. Chem. Soc.* **1948** : 2303.

BORSCHE, W., WAGNER-ROEMMICH, M., and BARTHENHEIER, J. (1942).—*Liebigs Ann.* **550** : 160.

ELIASBERG, J., and FRIEDLANDER, P. (1892).—*Ber. dtsch. chem. Ges.* **25** : 1752.

FIESER, L. F., and HERSHBERG, E. B. (1938).—*J. Amer. Chem. Soc.* **60** : 1893, 2542.

HUGHES, G. K. (1949).—*Aust. J. Sci. Res.* A **2** : 429.

HUGHES, G. K. (1950).—*Aust. J. Sci. Res.* A **3** : 497.

LAHEY, F. N. (1949).—*Aust. J. Sci. Res.* A **2** : 423, 622, 630.

LEHMSTEDT, K., and HUNDERTMARK, H. (1931).—*Ber. dtsch. chem. Ges.* **64** : 2386.

MERZ, J. H., and WATERS, W. A. (1949).—*J. Chem. Soc.* **1949** : 2427.

NIEMENTOWSKI, S. (1906).—*Ber. dtsch. chem. Ges.* **39** : 385.

POSTERNAK, T. (1936).—*Helv. Chim. Acta.* **19** : 1333.

PRICE, J. R. (1949).—*Aust. J. Sci. Res.* A **2** : 249, 255, 264, 272, 282.

RAO, K. V., SESHADRI, T. R., and THIRUVENGADAM, T. R. (1949).—*Proc. Indian Acad. Sci.* **30** : 114.

SCHÖPF, C., and LEHMANN, G. (1932).—*Liebigs Ann.* **497** : 7.

SHORT COMMUNICATIONS

A CIRCUIT FOR THE OPERATION OF A SYNCHROTRON IN CONJUNCTION WITH A WILSON CLOUD CHAMBER*

By E. G. Muirhead† and H. Lichtblau†

The output of betatron and synchrotron accelerators consists of a regular succession of pulses of a few microseconds duration. This fact is generally of little consequence in applications involving integrating exposures of detecting apparatus. The use of nuclear emulsions to detect disintegration products from nuclei and the induction of artificial radioactivity are the most important examples of this group. With the Wilson cloud chamber, however, it is essential to limit the irradiation of the cloud chamber to one or two output pulses which are timed to occur during the sensitive period following expansion of the chamber. The intense ionization produced by an accelerator, which is operated to give continuous output, paralyses the chamber and prevents the formation of tracks which are separately distinguishable.

The following account deals with a circuit allowing control of the number and timing of the X-ray pulses from a 14 MeV. electron synchrotron when operated in conjunction with a rubber diaphragm-type cloud chamber which has been described previously (Darby and Swan 1948). The chamber automatic control mechanism causes expansions to occur approximately once per minute and a timing signal for the circuit is derived at the completion of the movement of the expansion valve. Thus the synchronizing action is started at the beginning of the sensitive time, regardless of variations in the time taken for the mechanical motion. This signal gives rise to a gate pulse of variable duration which allows one of the periodically occurring electron injection trigger pulses to pass on and fire the injection modulator. After the time taken for acceleration, about eight milliseconds, an X-ray pulse is produced which occurs well within the sensitive time of most cloud chambers.

This method has the advantage of causing the least possible interference with the independent operation of the accelerator or cloud chamber. As it requires the continuous operation of the magnetic field of the accelerator, it might be undesirable for large accelerators, for which the amount of power dissipated between expansions would prove uneconomical. However, a similar method of synchronization has been successfully employed by Gaerttner and Yeater (1949) for a 70 MeV. synchrotron used in conjunction with a fast cycling expansion chamber.

---

* Manuscript received April 5, 1951.

† Physics Department, University of Melbourne.

The Circuit and Its Action

The downward motion of the chamber valve assembly closes the contacts of a microswitch (S_1) which causes the thyratron in the A.C. synchronizing circuit (Chute 1943, V_1 in Fig. 1) to fire. This circuit is energized from the same A.C. supply which delivers power to the magnetic field of the accelerator and its purpose is to synchronize the subsequent timing sequence relative to the electron injection phase in the vicinity of zero magnetic induction. A small adjustable delay governed by the R–C values in the control grid circuit of V_1 allows fine timing control, if desired. As long as the microswitch is depressed, V_1 will conduct on alternate half-cycles and a train of positive timing pulses is delivered from the differentiating circuit connected to the cathode of V_1. Passing to the next stage, the first timing pulse will discharge condenser C_5. Subsequent pulses are blocked by the previous interruption of the high tension supply to

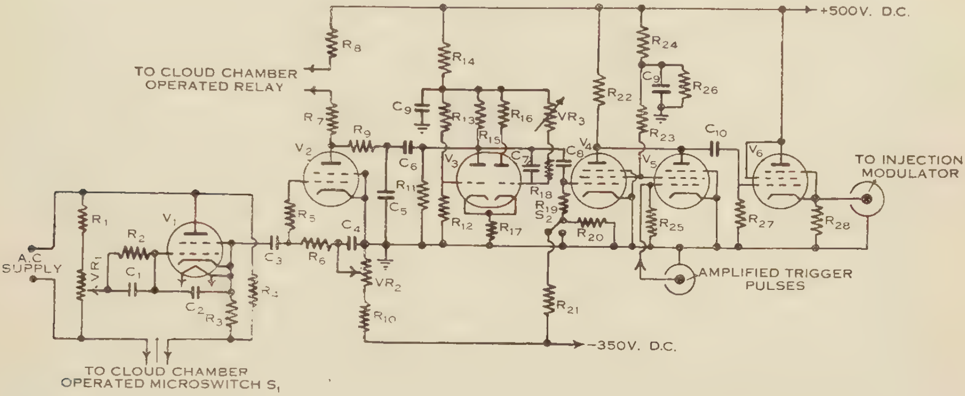

+500V. D.C.
R8
R14
R24
TO CLOUD CHAMBER OPERATED RELAY
C9
R26
R22
R7
R9
C9
R13
R16
VR3
R15
C7
C8
V4
R23
C10
V6
V2
C6
V3
V5
TO INJECTION MODULATOR
R11
C5
R18
R19
S2
R1
V1
R5
C4
R12
R17
R20
R25
R27
R28
A.C SUPPLY
R2
C3
R6
VR1
C1
VR2
AMPLIFIED TRIGGER PULSES
R4
C2
R3
R21
R10
−350V. D.C.
TO CLOUD CHAMBER OPERATED MICROSWITCH S1

Fig. 1.—Schematic circuit diagram.

Valves

V_1, V_2	2050	V_4, V_5, V_6	6AC7
V_3	6SN7–GT		

Resistors

R_1, R_2, R_7, R_8, R_{11}, R_{21}, R_{27}	500,000 Ω	R_{14}, R_{15}, R_{28}	20,000 Ω
R_3	3,000 Ω	R_{16}, R_{17}, R_{23}	10,000 Ω
R_4, R_6	50,000 Ω	R_{19}, R_{20}, R_{25}	100,000Ω
R_5	5,000 Ω	R_{22}	250,000 Ω
R_9	50 Ω	R_{24}	60,000 Ω
R_{10}, R_{18}	1 MΩ	R_{26}	150,000 Ω
R_{12}	25,000 Ω	VR_1	250,000 Ω
R_{13}	75,000 Ω	VR_2, VR_3	1 MΩ

Condensers

C_1	2 μF.	C_5	0·1 μF.
C_2	0·005 μF.	C_7	0·2 μF.
C_3, C_6	0·0001 μF.	C_8	0·5 μF.
C_4, C_9	8 μF.	C_{10}	0·02 μF.

V_2, conveniently effected by means of a relay operating from the chamber control circuits. The output pulse from V_2 triggers a multivibrator (V_3) which responds by forming a negative gate pulse of adjustable width. In this way, the motion of the valve assembly gives rise to a synchronized gate pulse at the output from V_3.

In order to time the injection process, the trigger pulses are derived from a small peaker strip in the accelerator magnet fringing field. After sufficient amplification, each one of these pulses is made to cut off one valve (V_4) of a Rossi coincidence circuit. The synchronized gate pulse when applied to the second Rossi valve (V_5) produces the desired coincidence output pulse which then proceeds to trigger the injection modulator via a cathode follower stage (V_6). The width of the gate pulse is usually adjusted to allow only one coincidence pulse to occur. The modulator incorporates the required amount of discrimination.

Operation

In order to put the combined system into operation, it is first necessary to run the accelerator on normal continuous operation during which time the various controls can be adjusted to their optimum settings for maximum output. For this purpose, the switch (S_2) allows the application of a fixed bias to cut off V_3, and so every trigger pulse will fire the modulator, and the accelerator works at the normal rate. After the desired operating conditions have been achieved, S_2 is opened and the circuit is ready for single pulse operation. Bias of V_2 is adjusted to cause the thyratron to fire on the steepest portions of the timing pulses. Finally, by means of a manually operated microswitch which simulates the motion of the cloud chamber, the output pulse is again adjusted to a maximum by altering slightly the injection phase control. A monitor ionization chamber recording the intensity of each X-ray pulse as well as a triggered high speed cathode ray oscillograph are almost indispensable for this purpose. By means of the oscillograph, it is possible to verify that the modulator fires once, and only once, while adjustments are made to the circuit prior to successful operation. Once operating, the circuit requires very little adjustment provided the magnetic field conditions are substantially unaltered. An occasional alteration of the injection phase setting, say once an hour in the present instance, is all that is necessary.

The authors gratefully acknowledge Professor L. H. Martin's continued encouragement and interest in their work. The assistance of Dr. W. B. Lasich was greatly appreciated.

References

Chute, G. M. (1943).—" Electronic Control of Resistance Welding." (McGraw-Hill Book Co. Inc.: New York.)

Darby, J. F., and Swan, J. B. (1948).—*Aust. J. Sci. Res.* A **1**: 18.

Gaerttner, E. R., and Yeater, M. L. (1949).—*Rev. Sci. Instrum.* **20**: 588.

SMALL SCALE PREPARATION OF METHYL RED INDICATOR*

By R. G. Gillis†

In the preparation of methyl red indicator, both the diazotization of anthranilic acid and the coupling with dimethylaniline are usually carried out in an aqueous medium, e.g. as in the procedure described by Clarke and Kirner (1932). Even on a smaller scale than they describe, the operations are time consuming and the purification tedious. In more than one laboratory, between five and ten grams are sufficient for many months' work ; the method described below is suitable for such small quantities, requires only a brief working time, and gives satisfactory yields. The essential difference of the method is that both diazotization (with butyl nitrite) and coupling are carried out in glacial acetic acid from which the crude indicator is precipitated free from inorganic salts.

Experimental

Anthranilic acid (6·85 g. ; 0·05 mole) is dissolved in glacial acetic acid (50 ml.) with warming if necessary. The solution is cooled in an ice-bath with stirring until the acetic acid begins to crystallize. *n*-Butyl nitrite (5·7 ml. ; 0·05 mole) (*Note 1*) is then added with stirring at such a rate that the temperature does not rise. When the addition is complete, the solution is allowed to stand for 15 minutes. Dimethylaniline (10·3 ml. ; 0·075 mole) is then added rapidly, the reaction mixture is stirred well and allowed to stand in ice and water for at least 3 hours, but preferably overnight. The precipitated methyl red is filtered, washed with 10 ml. of glacial acetic acid, and dried in a vacuum desiccator over solid sodium hydroxide. The crude yield is 7–8 g. (50–60%).

The crude material is best purified by Clarke and Kirner's method (loc. cit.) of extraction with boiling toluene using 10 ml. per gram of dry solid (*Note 2*). When the toluene extract is colourless, the flask is transferred to a boiling water-bath and allowed to cool slowly. After filtration and drying in a steam oven, the purified material appears as shining permanganate coloured plates, m.p. 178–179 °C. (uncorr.), and is suitable for most purposes ; if desired the purification by toluene extraction may be repeated. The efficiency of recovery in the purification stage is 80–90%.

Note 1.—*n*-Butyl nitrite is best prepared by the method of Noyes (1943).

Note 2.—The extraction is conveniently carried out in a " Quickfit " hot extraction apparatus (Catalogue No. EX 1/22). This holds a 25 by 80 mm. Soxhlet thimble which is suitable for a preparation on this scale.

References

Clarke, H. T., and Kirner, W. R. (1932).—" Organic Syntheses." Collective Vol. 1, p. 366. (Wiley : New York.)

Noyes, W. A. (1943).—" Organic Syntheses." Collective Vol. 2, p. 108. (Wiley : New York.)

---

* Manuscript received April 11, 1951.

† Department of Applied Chemistry, Melbourne Technical College.

CORRIGENDA

VOLUME 4, NUMBER 2

Page 148, Figure 9: *For* $\left(\frac{2\nu}{cx}\right)^{\frac{1}{2}}$ *read* $\left(\frac{2\nu}{cx}\right)^{\frac{1}{2}}\left(\frac{\omega}{c}\right)^{3/2}$

For $\left(\frac{2cx}{\nu}\right)^{\frac{1}{2}}$ *read* $\left(\frac{2cx}{\nu}\right)^{\frac{1}{2}}\left(\frac{\omega}{c}\right)^{\frac{1}{2}}$

Page 157, last line but one: *For* $\left(\frac{2cx}{\nu}\right)^{\frac{1}{2}}$ *read* $\left(\frac{2cx}{\nu}\right)^{\frac{1}{2}}\left(\frac{\omega}{c}\right)^{\frac{1}{2}}$

Page 157, last line: *For* $\left(\frac{2\nu x}{c}\right)^{\frac{1}{2}}$ *read* $\left(\frac{2\nu}{cx}\right)^{\frac{1}{2}}\left(\frac{\omega}{c}\right)^{3/2}$

COLLISION PROCESSES INVOLVING HIGHLY IONIZED ATOMS

By E. R. Hill*

[*Manuscript received June 27, 1951*]

Summary

The cross section of FeXIV for inelastic collisions resulting in the transition from the $^2P_{3/2}$ to the $^2P_{1/2}$ state of the ion, for electron collision ionization, and for radiative recombination of FeXIV and MgX are calculated by wave mechanical methods. A method is given whereby radiative recombination cross sections of all highly ionized atoms may be calculated. Where possible the cross sections obtained are compared with corresponding results given by the classical quantum theory. Finally, the wave mechanical values are used to verify the general conclusion, obtained using classical quantum values of the cross sections, that the agitation temperature of the coronal electrons is about 10^6 °K.

I. Introduction

It is now generally accepted that there exist within the solar corona a variety of highly ionized atoms (of which FeXIV and MgX are two of the most important), and electrons with an agitation temperature of the order of 10^6 °K. Furthermore, by using the fact that the equilibrium in the corona is known to be governed by the relative rates of radiative recombination and ionization by electron impact, Woolley (1947) was able to estimate the coronal electron temperature T. However, as a result of uncertainties in the cross sections used, it was thought worthwhile to attempt to improve the estimate of T using cross sections determined by wave mechanical methods. Consequently, cross sections have been calculated for ionization of FeXIV by electron impact, radiative recombination of FeXIV and MgX, and also collision excitation of FeXIV from the $^2P_{3/2}$ to $^2P_{1/2}$ state. The latter process is important because the radiative transition $^2P_{1/2}$ to $^2P_{3/2}$ gives rise to the prominent 5303 Å line observed in the coronal emission spectrum. The calculated cross sections are averaged over the Maxwell distribution of energies of the electrons in the solar corona so enabling direct comparison with experiment.

The inelastic collision cross section is evaluated by a procedure based on the method of distorted waves in which an approximate field of the ion, calculated using Slater's rules, is used to obtain accurate wave functions for the incident and scattered electrons. The averaged cross section obtained in this way is found to compare favourably with the experimental value obtained by Woolley and Allen (1949).

The ionization cross section is calculated using Born's approximation, and representing the ejected electron by a Coulomb wave function. The maximum ionization cross section of different atoms obtained by various authors using

---

* Physics Department, University of Melbourne.

classical quantum and wave mechanical theories are compared, and it is found that the classical quantum theory gives values which generally do not differ considerably from the wave mechanical cross sections.

Results obtained by various authors for the recombination of an electron with a proton are extended and used to obtain cross sections for radiative recombination of highly ionized atoms. Furthermore, by using the values of the cross sections for radiative recombination and collision ionization of FeXIV, it is verified that the electron agitation temperature of the solar corona is about 10^6 °K.

II. An Inelastic Collision Cross Section of FeXIV

The cross section of FeXIV for an electron collision which produces the transition $^2P_{3/2}\,(=A) \rightarrow {}^2P_{1/2}\,(=B)$ in the ion may, according to Mott and Massey (1949), be written approximately as

$$Q(A,k \rightarrow B,k')=\frac{4\pi^2 m^2}{h^4}\,\frac{k'}{k}\cdot\frac{1}{2(2J+1)}\sum_{\substack{ms,\, m's\\ M_J,\, M'_{J'}}}\int|\,(A,k\mid V\mid B,k')\,|^2 d\Omega, \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ (1)$$

where $k(h/2\pi)$ and $k'(h/2\pi)$ are the momenta of the incident and scattered electrons, $m_s(h/2\pi)$ and $m'_s(h/2\pi)$ are their respective spin components, $M_J(h/2\pi)$, and $M'_J/(h/2\pi)$ the allowed initial and final values of the total angular momentum of the ion, while $d\Omega$ is the element of solid angle into which the incident electron may be scattered. Furthermore, the matrix element $(A,k\mid V\mid B,k')$ is defined by

$$(A,k\mid V\mid B,k')=\int\Psi^*_{A,k}(\mathbf{r}_1,\mathbf{r}_2)V(\mathbf{r}_1,\mathbf{r}_2)\Psi_{B,k}'(\mathbf{r}_1,\mathbf{r}_2)d\tau_1 d\tau_2, \qquad \ldots\ldots\ (2)$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are respectively the position vectors of the incident and atomic electron, it being assumed that the ion may be considered as a one electron system. The wave functions $\psi_{A,k}$ and $\psi_{B,k}'$ represent the complete system (incident electron + ion) before and after the collision, while $V(\mathbf{r}_1,\mathbf{r}_2)$ was taken to be the potential energy of the scattered electron in the field of the ion, that is,

$$V(\mathbf{r}_1,\mathbf{r}_2)=\varepsilon^2\left(\frac{1}{r_{12}}-\frac{1}{r}\right),$$

where r is r_2 or r_1 according to whether electron exchange does or does not occur. Some workers have taken r to be r_1 for electron exchange, and recently Bates *et al.* (1950) have shown that neither is correct, and that one cannot at present hope to obtain better than the order of magnitude of the collision cross section in such cases.

The ψ-functions must be antisymmetrical in the coordinates of the two electrons so that we may write

$$\left.\begin{aligned}\Psi_{A,k}(\mathbf{r}_1,\mathbf{r}_2)&=2^{-\frac{1}{2}}[\psi_A(r_1)\sigma_1\chi_k(r_2)\sigma_2'-\psi_A(r_2)\sigma_2\chi_k(r_1)\sigma_1'],\\ \Psi_{B,k}'(\mathbf{r}_1,\mathbf{r}_2)&=2^{-\frac{1}{2}}[\psi_B(r_1)\varepsilon_1\chi'_k{}'(r_2)\varepsilon_2'-\psi_B(r_2)\varepsilon_2\chi'_k{}'(r_1)\varepsilon_1'],\end{aligned}\right\} \qquad \ldots\ldots\ (3)$$

where ψ_A and ψ_B are one electron wave functions representing the initial and final states of the ion, χ_k and $\chi'_k{}'$ are respectively the wave functions of the incident and scattered electrons, while σ and ε are the spin functions which may take either of the values α or β.

As the Ψ-functions represent states whose energy difference arises mainly from spin orbital interaction, it is necessary to use the proper linear combinations of hydrogen-like space wave functions and Pauli spin functions as given by Condon and Shortley (1935) for the initial and final state of the ion. In the hydrogen-like wave functions Z were replaced by $16 \cdot 5$ because $+16 \cdot 5\varepsilon$ is the effective nuclear charge in the vicinity of the $3p$-electron of FeXIV as determined from rules formulated by Slater (1930).

The wave function χ_k of the free incident electron in a field in which its potential energy $V(r)$ may be written as

$$\chi_k = k^{-\frac{1}{2}} r^{-1} \sum_{l=0}^{\infty} L_l(k,r) P_l(\cos\theta) \begin{Bmatrix} \alpha(s) \\ \beta(s) \end{Bmatrix}, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

where

$$L_l(k,r) = (2l+1) i^l e^{i\eta_l} G_l(r),$$

and where G_l is the solution of the differential equation

$$G_l'' + \left[k^2 - V(r) - \frac{l(l+1)}{r^2}\right] G_l = 0, \quad \ldots\ldots\ldots \quad (5)$$

which behaves like r^l for small r and asymptotes to $k^{-\frac{1}{2}} \sin(kr + \eta_l)$, η_l being the phase shift whose actual magnitude is unimportant for this calculation. Thus in order to calculate the continuous wave function it is necessary to know the potential energy $V(r)$ for an electron at any point in the field of FeXIV, and since no details concerning $V(r)$ existed at the time of commencing this work it was necessary to calculate an approximate field using Slater's rules. These rules give the effective nuclear charge in the vicinity of each electron shell and hence for all radial distances by graphical interpolation and extrapolation onto a Coulomb field $+13\varepsilon/r$ at large radial distances.* Both the calculated potential and the Coulomb potential are shown in Figure 1.

In the same diagram the form of the normalized distorted wave G_0 is shown for the field of FeXIV and for the Coulomb field.† It will be noticed that the strong attractive field of the ion pulls the electron wave in towards the nucleus so strongly that the wavelength is about $0 \cdot 14 a_0$, whereas the corresponding length of the plane wave representing an incident electron of the same energy is $6 \cdot 3 a_0$. Furthermore, the distorted wave has a much smaller amplitude than the corresponding plane wave and asymptotes to the latter only at extremely large radial distances. As a result of the smaller amplitude and wavelength one must expect that the distorted waves will give a much smaller inelastic cross section than the simpler plane wave approximation. It is also clear that the same remarks apply to a lesser extent, to the differences between the Coulomb wave and the distorted wave.

---

* It will be noted at this stage that the author's method differs from that of Hebb and Menzel (1940) in that the latter used a Coulomb potential to represent the field of O III for all radial distances.

† The normalization process was carried out in the manner described by Bates *et al.* (1939).

The scattered electron wave function $\chi'_{k'}$ may be represented in a similar manner to that of the incident electron by

$$\chi'_{k'}=r^{-1}k'^{-\frac{1}{2}}\sum_{l'=0}^{\infty}L_{l'}(k',r')P_{l'}(\cos\Theta), \quad \ldots\ldots\ldots\ldots \quad (6)$$

where Θ is the angle between the direction of motion of the scattered electron and the radius vector $\mathbf{r}$ to the point in space at which χ' is represented. $L_{l}'(k',r')$ is the function defined in equation (4) with l, k, and r replaced by l', k', and r'. By using the addition theorem for the Legendre polynomial (see Copson 1935), $P(\cos\Theta)$ is expressed in terms of Legendre polynomials involving the angles θ and θ' which the incident electron path makes with the direction of motion of the scattered electron and with the radius vector $\mathbf{r}$. Since the transition energy (2·33 eV.) is small compared with the incident electron energies which are of interest in these calculations, we may put $k=k'$. In any case χ has been so defined that the G_l-functions are influenced only slightly by variations in k and k'.

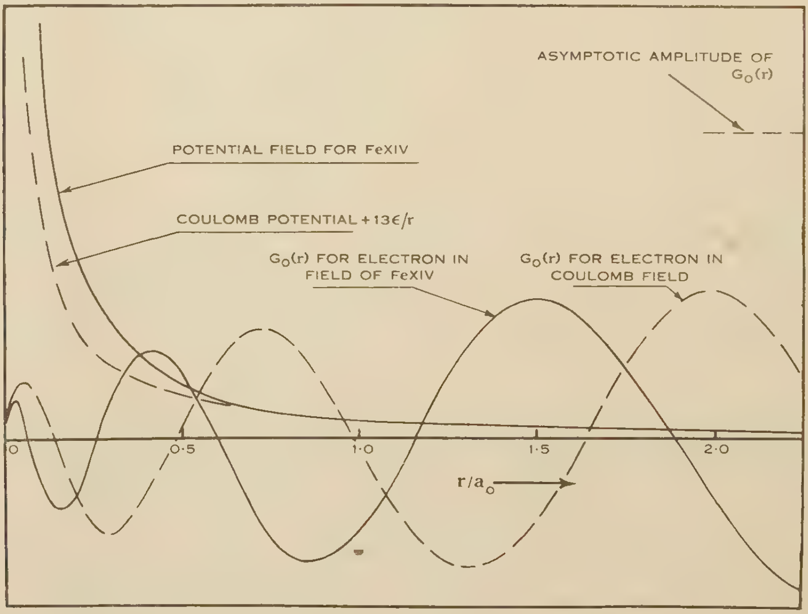


Fig. 1.—The potential field of FeXIV as calculated from Slater's rules, and for comparison, the Coulomb potential field for radial distances up to $2\cdot5a_0$ from the nucleus. The normalized G_0-functions together with their asymptotic amplitude are shown for an electron incident upon the abovementioned fields with an energy of 3·8 eV.

The cross section for transitions between each of the degenerate levels may now be calculated using antisymmetrical combinations (4) of wave functions of the FeXIV electron and free electron, and averaging over all the possible spin orientations of the two electrons. For each initial electron spin array this averaging will introduce a factor $\frac{1}{4}$, so that the total cross section for the transitions between any pair of the degenerate levels is found by adding the cross sections obtained for the various spin orientations. The total cross section for transition between states A and B is then found by adding all the transition cross sections for the individual multiplet levels in each state and dividing by 2 since it is equally likely for the FeXIV electron that m_s is $+\frac{1}{2}$ or $-\frac{1}{2}$ initially.

Finally, as the radial matrix elements do not vary significantly with incident electron energy because of the near independence of the G_l's with respect to k it is found that the cross section may be written in the form $2\cdot 22\times 10^7\ v^{-2}\pi a_0^2$. When considering electron collisions in the solar corona where the electron distribution is Maxwellian, the collision cross section must be averaged over this distribution; the averaged cross section is found to be $0\cdot 11\pi a_0^2$.

This value is to be compared with the experimental value of $0\cdot 26\pi a_0^2$ obtained by Woolley and Allen (1949). In view of the difficulties of the experimental determination as well as the inaccuracy of the distorted wave method for the energies involved, this agreement is satisfactory. In addition it is found that the cross section satisfies the conservation theorems (see Mott and Massey 1949), which indicate upper limits for inelastic collision cross sections. However, the cross sections obtained by Hebb and Menzel (1940) using the distorted wave method violate the theorems. This is most probably due to the fact that they used Coulomb G-functions corresponding to an effective nuclear charge $+2\varepsilon$ which is too small in the vicinity of the electron shells. This has the effect of decreasing the number of oscillations in the G-functions, so that there is increased cancellation due to interference of the matrix elements.

III. Ionization Collision Cross Section of FeXIV

The cross section $Q_I(k^2)$ of an atomic system for electron collision ionization is given by the Born approximation as

$$Q_I(k^2)=\frac{4\pi^2m^2}{h^4k}\int k'\,|\,(A,k\,|\,V\,|\,\varkappa,k')\,|^2\,d\omega d\Omega d\varkappa, \quad \ldots\ldots\ldots \quad (7)$$

where $d\Omega$, k, and k' are as defined in Section II, $d\omega$ is the element of solid angle into which the atomic electron is ejected, and $\varkappa(h/2\pi)$ is the momentum of the ejected electron. The wave numbers of the incident, scattered, and ejected electrons are related to the ionization energy E_I by

$$k^2=k'^2+\varkappa^2+8\pi^2mE_I/h^2, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (8)$$

where E_I is 355 eV. for the valence electron of FeXIV.

As before, $(A,k\mid V\mid \varkappa,k')$ is the matrix element of the interaction between the incident electron and the atomic system. But for simplicity no allowance is made for electron exchange or the distortion, by the scattering field, of the incident and scattered electron wave function. Consequently the wave function of the system before and after the collision is written as the product of the appropriate electron wave functions; plane waves being used to represent the incident and scattered electrons and a Coulomb wave function representing the ejected electron. Since it is found that the matrix element may be evaluated most conveniently using parabolic coordinates, it is usual to write the ejected electron wave function in the form due to Sommerfeld (1931). The effective nuclear charge appearing in this expression is, according to Massey (see Burhop 1940), the value $+16\cdot5\varepsilon$ obtained using Slater's rules. The Coulomb wave function chosen above has been used by Massey and Mohr (1933) to allow for the large interaction between the ion and slow moving ejected electron because it is this type of collision which provides the main contribution to the ionization cross section.

To reduce the computational work a hydrogen-like wave function is used for the atomic electron and considerable complication may be removed by using a $3s$-hydrogen-like wave function instead of the $3p$-function. The effect of the alteration on the calculated cross section should be small; thus Burhop (1940) found that the ionization cross section per electron was practically the same for the ${}^2s_{1/2}$, ${}^2p_{1/2}$, and ${}^2p_{3/2}$ electrons in Ag and Hg. On substituting the above wave functions in expression (7), integrating over the coordinates of the incident electron using Bethe's integral, and then changing to parabolic coordinates ξ, η, and φ we get

$$2\pi N \int \exp(\xi A-\eta B-u)u^{-in}J_0\{2C\xi\}J_0\{2D\eta^{\frac{1}{2}}\} \times\left[\frac{\mu^2}{6}(\xi+\eta)^3-\mu(\xi+\eta)^2+(\xi+\eta)\right]du\,d\xi\,d\eta, \qquad \ldots\ldots (9)$$

where

$$A=\tfrac{1}{2}(\mu-i\varkappa-iK), \quad C=(i\varkappa u\cos^{\frac{1}{2}}\theta/2)^{\frac{1}{2}}, \quad \mu=Z/3=n\varkappa a_0/3,$$
$$B=\tfrac{1}{2}(\mu-i\varkappa+iK), \quad D=(i\varkappa u\sin^2\theta/2)^{\frac{1}{2}}, \quad K^2=k^2+k'^2-2kk'\cos\delta$$

(δ=scattering angle),

and

$$N=\frac{64\pi m^2Z^3}{27kh^4}\left(\frac{n}{1-e^{-2\pi n}}\right)^{\frac{1}{2}}\frac{1}{|\Gamma(1-in)|^2}.$$

The integrations over ξ and η may be carried out with the aid of the integral

$$\int_0^\infty e^{-\lambda x}x^{n}J_0(2r\xi^{\frac{1}{2}})dx=\frac{n!}{\lambda^{n+1}}e^{-r^2/\lambda}L_n(r^2/\lambda), \qquad \ldots\ldots\ldots (10)$$

but as the result of these integrations is extremely lengthy, it will not be written down. Although it is possible to perform the next integration analytically, the expression obtained was so long that it was decided to carry out the remaining integrations numerically.

Since the smaller incident electron energies contribute most when averaging over a Maxwellian distribution of incident electron energies, the calculations were carried out for an incident electron energy of 900 eV. where the cross section might be expected to be near its maximum value. Under these circumstances the velocity distribution of the ejected electrons will be nearly symmetrical about its midpoint corresponding to 136 eV., and because of the extreme length of the calculations it was decided to carry them through for one ejected electron energy only, namely 136 eV., which corresponds to the maximum of an assumed parabolic ejected electron velocity distribution.

On substituting numerical values into (9) and performing the necessary integration, the cross section per electron for ionization of FeXIV was found to be $1\cdot0\times10^{-19}$ cm.^{2} for an incident electron energy of 900 eV. The cross section for other incident electron energies was then obtained by taking a mean of all ionization functions resulting from similar calculations of which the author was aware at the time of commencing this work. The dispersion of these ionization functions is illustrated by Figure 2, curves 1 and 2, which represent the upper and lower limits of the relative cross sections obtained by Massey and Mohr (1933) for H and He, and Burhop (1940) for K-shell ionization of Ag, Hg, and Ni, and L-shell ionization of Ag and Hg. Curves 3 and 4 are the corresponding experimental curves determined from the results of Tate and Smith (1932, 1934) for H and He, Pockman *et al.* (1947) for Ni(K), Webster, Hansen, and Duveneck (1933) for Ag(K), and McCue (1944) for Ag(L). A mean of curves 1 and 2 was readily fixed because of the remarkable uniformity of these functions and this mean curve was used in the following calculations. Curve 5 shows the ionization function obtained by Yavorsky (1946) and F. C. Ledsham (unpublished data) for Hg and Ne, while curves 6 and 7 indicate the dispersion of the corresponding experimental results of Nottingham (1939)

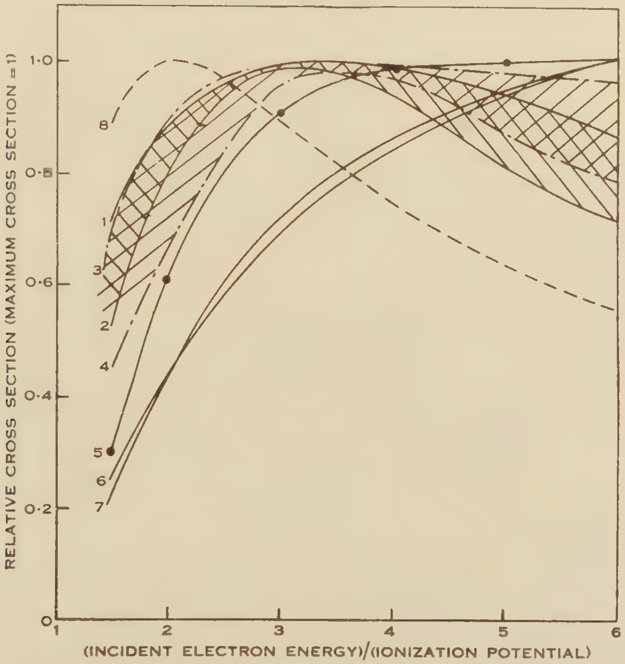


Fig. 2.—The dispersion of theoretical and experimental relative cross section curves. Curves 1 and 2 respectively indicate the maximum and minimum values of the relative cross section used to determine the ionization function of FeXIV, while curves 3 and 4 are the corresponding experimental curves for atoms concerned in curves 1 and 2. The theoretical curve 5 and the corresponding experimental curves 6 and 7 relate to theoretical work of which the author became aware after the completion of the present calculation. Finally, curve 8 shows the result obtained using the classical theory.

and Bleakney (1930). On averaging the ionization function over the Maxwell energy distribution of the coronal electrons a mean cross section $1\cdot0\times10^{-19}$ cm. was obtained.

For practical purposes it is of considerable interest to compare ionization cross sections given by the classical and wave mechanical theories. Thomson (1912) has shown that the ionization cross section, according to the classical theory, is given by

$$Q_I(k^2)=\frac{8\pi^2 m\varepsilon^4}{k^2h^2}\left(\frac{1}{E_I}-\frac{8\pi^2 m}{k^2h^2}\right), \quad \ldots\ldots\ldots\ldots (11)$$

so that when $\frac{k^2h^2}{8\pi^2m}=2E_I, Q_I(k^2)$ has its maximum of

$$\pi\varepsilon^2/4E_I^2. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (12)$$

Table 1 contains the maximum cross sections given by wave mechanical theory and those obtained from expression (**12**) using experimental measurements of E_I quoted in the second column, and both are compared with the experimental results obtained for the ionization cross sections of different electron shells of various elements.

TABLE 1

COMPARISON OF IONIZATION CROSS SECTIONS

Atom or Molecule and Shell Ionized	Experimental Ionization Potential (eV.)	Maximum Cross Section per Electron		
		Experimental (cm.^{2})	Quantum Mechanical (cm.^{2})	Classical (cm.^{2})
$H_2(K)$	$15\cdot9$	$0\cdot5\times10^{-16}$	$0\cdot76\times10^{-16}$	$0\cdot64\times10^{-16}$
He(K)	$24\cdot5$	$0\cdot162\times10^{-16}$	$0\cdot232\times10^{-16}$	$0\cdot274\times10^{-16}$
Ni(K)	8300	$2\cdot1\times10^{-22}$	$2\cdot1\times10^{-22}$	$2\cdot37\times10^{-22}$
Ag(K)	24,900	$2\cdot5\times10^{-23}$	$2\cdot46\times10^{-23}$	$2\cdot64\times10^{-23}$
Hg(K)	82,000	—	$2\cdot66\times10^{-24}$	$2\cdot39\times10^{-24}$
Ne(L)	$21\cdot56$	$0\cdot724\times10^{-16}$	$1\cdot04\times10^{-16}$	$0\cdot353\times10^{-16}$
Ag($L_{\rm I}$)	3,800	—	$7\cdot06\times10^{-22}$	$11\cdot3\times10^{-22}$
Ag($L_{\rm II}$)	3,540	—	$8\cdot30\times10^{-22}$	$13\cdot8\times10^{-22}$
Ag($L_{\rm III}$)	3,340	—	$8\cdot50\times10^{-22}$	$14\cdot7\times10^{-22}$
Hg($L_{\rm I}$)	14,720	—	$6\cdot13\times10^{-23}$	$7\cdot57\times10^{-23}$
Hg($L_{\rm II}$)	14,130	—	$6\cdot39\times10^{-23}$	$8\cdot21\times10^{-23}$
Hg($L_{\rm III}$)	12,300	—	$7\cdot25\times10^{-23}$	$11\cdot0\times10^{-23}$
FeXIV($M_{\rm II}$) ..	355	—	$1\cdot0\times10^{-19}$	$1\cdot36\times10^{-19}$

Table 1 shows how well the classical theory estimates the order of magnitude of the maximum ionization cross section, thus indicating that the results of this theory may be used in many cases to get a first approximation to the rates of electron collision ionization reactions. In more accurate work it would be desirable to combine the cross sections obtained in this way with an ionization function similar to that obtained from Figure 2.

IV. RADIATIVE RECOMBINATION OF ELECTRONS WITH FeXIV IONS

The probability of electron capture for various types of atomic systems such as positive and negative ions as well as neutral atoms is calculated by wave mechanical methods in terms of cross sections which we shall denote by Q_R^t. Modern quantum theory gives the following expression for the cross section of an atomic system for radiative recombination,

$$Q_R^t(k^2)=\sum_s Q_r^s(k^2)=\frac{128\pi^5 m\varepsilon^2}{3h^2c^3k}\sum_s \nu_{ks}\,\left|\int \psi_s{}^*\mathbf{r}\psi_k d\tau\right|^2, \quad \ldots\ldots \quad (13)$$

where ψ_s is the wave function for the bound state E_s into which the electron falls, ψ_k is the continuous wave function representing the incident electron while in its positive energy state prior to capture, and $\mathbf{r}$ is its vector position coordinate. The frequency ν_{ks} of the emitted radiation is related to the final bound state energy by the Bohr frequency condition, so that

$$\frac{k^2h^2}{8\pi^2 m}-E_s=h\nu_{ks}. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (14)$$

The summation in (**13**) is over all the unoccupied states of the capturing system into which the incident electron may fall.

Apart from the simple case of recombination of an electron to a proton for which (**13**) may be evaluated analytically (see Wessel 1930), numerical methods using wave functions calculated by the Hartree self-consistent field method have to be employed. Bates (1946) has shown that except for electron attachment to excited states, the hydrogen-like approximation cannot be used generally because of the inequality of the effective nuclear charges appropriate to the bound and continuous wave functions. However, in the case of recombination to a highly ionized atom such as FeXIV, or MgX, the hydrogen-like approximation will be quite satisfactory for calculating the probability for capture to the ground state as well as to the excited states.* It is likely that the error in the cross section for capture to the ground state will be greater than to the excited states, but this will not be significant because of the predominant contribution to the total capture cross section by the latter processes.

In the solar corona one finds atoms which have had as many as 10 to 14 electrons stripped from them (Hoyle 1949), and to facilitate the estimation of recombination coefficients for such ions, we shall systematize and extend the results available with the use of the hydrogen-like approximation.

Wessel (1930) has evaluated the matrix elements of (**13**) in spherical polar coordinates, and obtained analytic formulae for the cross sections of electron capture to the ground state and the first two excited states denoted by the quantum number n and azimuthal number l. He has shown that

$$Q_R^{nl}(k^2)=\frac{64h\varepsilon^2}{3m^2c^3}\left(\frac{1}{ka}\right)^2 e^{-4/ka\ \mathrm{artan}\ nka}(1-e^{-2\pi/ka})f(n,l,1/ka), \quad \ldots\ldots\ldots \quad (15)$$

---

* This is because the binding of the remaining electrons is so tight that the effective nuclear charge over the whole range for which considerable contributions to the integral in (**13**) will arise, is equal to the atomic number minus the number of remaining electrons.

where the f's are simple expressions presented in Table 2, while

$$(ka)^2 = 0{\cdot}074E/Z^2 = p, \quad a = a_0/Z = h^2/4\pi^2 m\varepsilon^2 Z, \quad \ldots\ldots\ldots \quad (16)$$

where E is the energy of the incident electron in electron volts and the effective nuclear charge of the capturing ion.

TABLE 2

TABULATION OF WESSEL'S f-FUNCTIONS FOR VARIOUS ELECTRON STATES*

Total Quantum Number (h)	Aximuthal Quantum Number (l)	$f(n, l, 1/ka)$
1	0	$Z^4(Z^2+p)^{-2}$
2	0	$8Z^4(Z^2+p)^2(Z^2+4p)^{-3}$
	1	$2Z^6(Z^2+p)^2(Z^2+4p)^{-4}$
3	0	$3Z^4(Z^2+p)^2(Z^2+27p)^2(Z^2+9p)^{-5}$
	1	$24(19Z^2+90pZ+81p^2)(Z^2+9p)^{-5}$
	2	$48(Z^2+p)(11Z^2+45p)(Z^2+9p)^{-6}$

* The slight differences in the powers and numeral coefficients between the above tabulations are due to misprints in Wessel's paper.

It is seen from the above expressions that Z and v always appear together as v/Z which is proportional to ka. Consequently, on determining the partial cross section for capture by a proton ($Z=1$) for various $(ka)^{-1}$, one may readily calculate the capture cross section of any system for which the hydrogen-like approximation is valid. The method used is based on equations (**15**) and (**16**) from which it follows that the cross section for capture, by a proton, of an electron with energy Z^2E electron volts is the same as that for capture of an electron volt with energy E electron volts by a hydrogen-like system in which the effective nuclear charge is $+Z\varepsilon$. Since the coronal ions with which we are interested differ only in the populations of shells containing electrons with $n=1$, 2, and 3 (total quantum number), all that is necessary for a calculation of their $Q^t_R(k^2)$ is a tabulation for various incident electron energies of $Q^n_R(k^2)$ of a proton for $n=1, 2, 3$, and the various substates associated with these quantum numbers as well as $\sum_n Q^n_R$ for all states corresponding to $n>3$.

The electron capture cross section to the three least excited states of a proton may be calculated readily from (**15**) in conjunction with Table 2 for any value of $(ka)^{-1}$. The total capture cross sections are graphed for $(ka)^{-1}$ up to 50 in Figure 3. That part of the graph lying between $7 \leqslant ka \leqslant 20$ is obtained from the work of Bates *et al.* (1939), whilst the remaining portions of the curves were obtained as follows. For $(ka)^{-1} < 7$ a graph of the ratio (Q^1_R/Q^t_R) obtained from the abovementioned work is extrapolated to the value 1 for $(ka)^{-1} = 0$, which is the limit of the ratio as found from Wessel's analytical relationships for the two cross sections. Then calculating Q_R^1 from (**15**) for various $(ka)^{-1}<7$ and using this in conjunction with the graph, a good estimate of Q_R^t for $(ka)^{-1}$ is obtained. The extrapolation for $(ka)^{-1}>20$ may be performed readily with the aid of Kramers's classical quantum expressions for Q_R^n. Kramers (1923) was able to show that

$$Q_R^n=\frac{512\pi^6\varepsilon^{10}m}{3^{3/2}h^6c^3k^2n^3\nu_{kn}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ (\mathbf{17})$$

so long as

$$\frac{16\pi^4\nu_{kn}\varepsilon^2m^2}{h^3k^3}\gg 1.$$

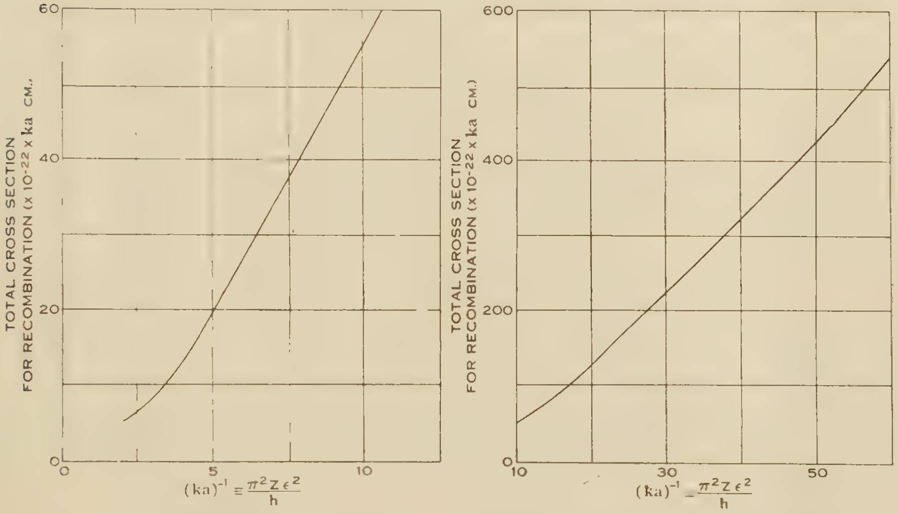


Fig. 3.—A plot against $(ka)^{-1}$ of the total capture cross section of a hydrogen atom as calculated by wave mechanical methods.

Since one may readily show that for $(ka)^{-1}>20$, the inequality is completely satisfied, we may safely use (**17**). Using the Bohr frequency condition to eliminate ν_{kn}, (**17**) may be shown to become

$$Q_R^n=\frac{16\varepsilon^2h}{3^{3/2}c^3n^2}\left(\frac{1}{ka}\right)^4\frac{1}{m(n^2+(ka)^{-2})}. \quad \ldots\ldots\ldots\ldots\ (\mathbf{18})$$

So that, providing that (**17**) is sufficiently accurate, the Q_R^t may be found on summing the series

$$\sum_{n=1}^{\infty}\frac{1}{n(n^2+(ka)^{-2})},$$

for various $(ka)^{-1}$. Comparison of (**17**) with the corresponding cross sections found from Wessel's results indicates that, except for the first few values of n, the differences are of no significance. The deviations for $n=1,2,3$ which never exceed more than 20 per cent. may be readily corrected by replacing the Kramers formula values by those found from (**15**). The necessity for such corrections becomes more important with increase in $(ka)^{-1}$.

Radiative recombination coefficients of FeXIV and MgX, averaged over the electron energy distribution in the solar corona, were computed using cross sections calculated by the method just described and were found to have the values $3 \cdot 4 \times 10^{-12}$ and $2 \cdot 0 \times 10^{-12}$ cm. sec. respectively. Woolley (1947) has estimated the average recombination coefficient of FeXIV using the results of Bates *et al.* (1939) for O II, assuming a Z dependence for the coefficient but making no allowance for the different electronic configurations of the two systems with the result that the value obtained was twice as large as that quoted above.

V. Application to the Solar Corona

The above wave mechanical estimates of the cross section of FeXIV for electron collision ionization and radiative recombination may be used to test the accuracy of previous estimates of the agitation temperature of the solar corona. If one supposes that the density of ions of a particular element (such as FeXIV) in two successive stages of ionization be $n(1-x)$ and nx (the latter ions being ionized more than the former), then on balancing the rates of collision ionization and radiative recombination governing transitions between these two stages of ionization, one may write the ratio of the two populations of these two stages as

$$\text{Degree of ionization} = \frac{x}{1-x} = \frac{(Q_I)_{av}}{(Q_R)_{av}}, \qquad \ldots\ldots\ldots (19)$$

where the suffix *av* indicates that cross sections have been averaged over the Maxwell distribution of the incident electron energies corresponding to a temperature T. For each particular element, the most abundant of its ions will be those for which $(x/1-x)=1$ (see Hoyle 1949), and it has been shown experimentally by Edlén (1942) that such ions have an ionization potential of about 400 eV.

Substituting the values of the relevant cross sections of FeXIV into equation (**19**) one finds that the experimental result is satisfied when T has the value $1 \cdot 1 \times 10^6$ °K., thus substantiating the generally accepted value of 10^6 °K. for the agitation temperature of the corona.

VI. Acknowledgments

In conclusion the author wishes to thank Professor R. v. d. R. Woolley for suggesting that this investigation be made ; and to acknowledge the constant advice and assistance of Associate Professor C. B. O. Mohr on all matters connected with this paper.

VII. References

Bates, D. R., Buckingham, R. A., Massey, H. S. W., and Unwin, J. J. (1939).—*Proc. Roy. Soc.* A **170** : 322.

Bates, D. R. (1946).—*Mon. Not. R. Astr. Soc.* **100** : 25.

Bates, D. R., Fundaminsky, A., Massey, H. S. W., and Leech, J. W. (1950).—*Philos. Trans.* A **243** : 93.

Bleakney, W. (1930).—*Phys. Rev.* **36** : 1303.

Burhop, E. H. S. (1940).—*Proc. Camb. Phil. Soc.* **30** : 43.

CONDON, E. U., and SHORTLEY, G. H. (1935).—" The Theory of Atomic Spectra." (Cambridge Univ. Press.)

COPSON, E. T. (1935).—" Theory of Functions of a Complex Variable." p. 304. (Oxford Univ. Press.)

EDLÉN, B. (1942).—*Ark. Mat. Astr. Fys.* **28** : 1.

HEBB, M. H., and MENZEL, D. H. (1940).—*Astrophys. J.* **92** : 408.

HOYLE, F. (1949).—" Some Recent Researches in Solar Physics." (Cambridge Univ. Press.)

KRAMERS, H. A. (1923).—*Phil. Mag.* **46** : 836.

MASSEY, H. S. W., and MOHR, C. B. O. (1933).—*Proc. Roy. Soc.* A **140** : 613.

McCUE, J. J. (1944).—*Phys. Rev.* **65** : 168.

MOTT, N. F., and MASSEY, H. S. W. (1949).—" The Theory of Atomic Collisions." 2nd Ed. (Oxford Univ. Press.)

NOTTINGHAM, W. B. (1939).—*Phys. Rev.* **55** : 203.

POCKMAN, L. T., WEBSTER, D. L., KIRKPATRICK, P., and HARWORTH, K. (1947).—*Phys. Rev.* **71** : 330.

SLATER, J. C. (1930).—*Phys. Rev.* **36** : 57.

SOMMERFELD, A. (1931).—*Ann. Phys. Lpz.* **11** : 257.

TATE, J. T., and SMITH, P. T. (1932).—*Phys. Rev.* **39** : 270.

TATE, J. T., and SMITH, P. T. (1934).—*Phys. Rev.* **46** : 773.

THOMSON, J. J. (1912).—*Phil. Mag.* **23** : 449.

WEBSTER, D. L., HANSEN, W. W., and DUVENECK, F. B. (1933).—*Phys. Rev.* **43** : 839.

WESSEL, W. (1930).—*Ann. Phys. Lpz.* **5** : 611.

WOOLLEY, R. v. d. R. (1947).—*Aust. J. Sci.* **10** : Suppl. No. 2.

WOOLLEY, R. v. d. R., and ALLEN, C. W. (1949).—*Mon. Not. R. Astr. Soc.* **108** : 293.

YAVORSKY, B. M. (1946).—*J. Phys. U.S.S.R.* **10** : 476.

THE STRAGGLING OF ELECTRONS BELOW THE CRITICAL ENERGY

By K. C. Hines*

[*Manuscript received May* 17, 1951]

Summary

The straggling in range of electrons which lose energy by collision and radiation processes is investigated. A diffusion equation is set up and a solution obtained following the method of Bhabha and Chakrabarty. In a given material the results apply for initial energies up to about half the critical energy. A discussion of the application of the results to the interpretation of experiments on meson decay electrons is given.

I. Introduction

In traversing absorbing materials made up of elements with low atomic number, electrons of energy between about 15 and 55 MeV. lose energy by ionization processes (collisions with bound electrons in the absorber) and radiation processes (involving deflection of the electrons in the Coulomb fields of nuclei and the consequent emission of photons). The energy loss by radiation is a result principally of a relatively small number of processes during which large fractions of the electron energy may be lost.

If a beam of mono-energetic electrons is incident on a layer of absorbing material, this characteristic mode of energy loss has the following effect. At a given penetration depth some electrons have undergone more radiation processes than others with the result that the electrons, instead of having a common range, display straggling.

In contrast to the radiation processes, energy loss by ionization takes place in many events in which only a small amount of energy is lost. Because of this the contribution of collision processes to straggling is very small compared with that of the radiation processes.

A further cause of straggling with electrons of these energies is the reduction in range resulting from the deflections in the Coulomb fields of nuclei, i.e. straggling produced by scattering.

The radiation straggling of electrons has been investigated by Corben (1938) and by Eyges (1949, 1950), both of whom used approximate expressions for the bremsstrahlung cross section formula, and neglected the pair production by emitted photons which could give rise to observable electrons at far greater penetration depths than the average primary electron range. Both these factors will introduce appreciable errors in the predicted ranges for the energies under consideration. Corben gives a solution which is valid for initial energies from two or three times the critical energy up to about ten times the critical

---

* Physics Department, University of Melbourne.

energy β, where β denotes the energy at which loss by radiation is about equal to the loss by collision ($\beta \sim 800$ MeV./Z). In the case of electrons with energies below the critical energy his series solution diverges. Eyges gives results for initial energies near and above the critical energy and for subsequent energies not too far below the initial energy.

The results of both these authors then, apply to the case of electrons which lose the greater part of their energy by radiation processes. The present work is an attempt to calculate the straggling in the case of electrons having initial energies below the critical energy. The calculations are of possible practical significance since the initial energies of meson decay electrons lie in this range. The electrons considered will lose the greater part of their energy by collision processes.

II. Calculation

For the investigation of the radiation straggling it is necessary to set up the equation for the probability of the electron appearing at any depth in the absorbing material, considering both collision and radiation processes.

The cross section for the emission by an electron of energy E of a quantum of energy between k and $k+dk$ while passing through the field of an atom with nuclear charge Ze is

$$d\varphi_k = \frac{4Z^2}{137}\left(\frac{e^2}{mc^2}\right)^2 \log 183Z^{-1/3} F(E,k)\frac{kdk}{E^2}, \quad \ldots\ldots\ldots \quad (1)$$

where e is the charge and m the mass of the electron and in the case of complete screening (high energies)

$$F(E,k) = 1 - \left(\frac{4}{3}+\alpha\right)\left(\frac{E}{k} - \frac{E^2}{k^2}\right). \quad \ldots\ldots\ldots\ldots \quad (2)$$

Because of the complexity of the exact formulae for incomplete screening, it has been necessary to use the expression (1) for the present calculations. This introduces an error of about 10 per cent. in the results.

For convenience we write (1) in the form

$$d\varphi_k = (\xi E, kdk)/l, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

with $$\frac{4Z^2}{137}\left(\frac{e^2}{mc^2}\right)^2 \log 183Z^{-1/3} = \frac{1}{l}\,; \quad F(E,k)\frac{kdk}{E^2} = \xi(E,k)dk. \quad \ldots\ldots \quad (4)$$

Thus the probability that an electron of energy E in travelling a distance dx cm. through a material with N atoms per unit volume will emit a quantum with energy between k and $k+dk$ is given by

$$\xi(E,k)dkNdx/l. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

It is convenient to introduce a new unit of length defined by

$$t = x/l. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (6)$$

Expression (5) now gives as the probability of emission in traversing a distance dt radiation units

$$\xi(E,k)dkNdt. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (7)$$

Using this unit of length the probability given by (7) is the same for all elements. The radiation unit is about 43 cm. for water, 9·8 cm. for aluminium, and 0·53 cm. for lead.

It is possible, to a good approximation, to take the collision loss as a constant energy loss per unit radiation length. Any error introduced by this assumption will affect only the mean range of the electrons, not the straggling in range, since this latter is a radiation effect. We denote the energy lost by collision processes per unit radiation length by β.

Suppose that a beam of electrons of initial energy E_0 is incident normally on a block of absorbing material. These electrons will lose energy by collision and radiation in the block and we denote by $h(E,t)dE$ the number of electrons with energy between E and $E+dE$ at a depth t in the material. The number of initial electrons will be represented by $h(E_0,0)$.

The number of electrons at depth t, which lose an energy between k and $k+dk$ by radiation in traversing an additional distance dt, is

$$\xi(E,k)h(E,t)dEdtdk. \quad \dots\dots\dots\dots\dots\dots \quad (8)$$

The number of electrons which suffer some radiation loss in dt is obtained by integrating (8) over all possible values of k, so that the number of electrons at depth t which suffer no radiation loss in dt is given by

$$h(E,t)dE-\int_0^E \xi(E,k)h(E,t)dEdtdk. \quad \dots\dots\dots \quad (9)$$

These electrons will all undergo loss of energy by collision of an amount βdt in traversing dt and will appear at $t+dt$ with an energy $E-\beta dt$.

The total number of electrons of energy between $E-\beta dt$ and $E+dE-\beta dt$ at $t+dt$ is

$$h(E-\beta dt,\, t+dt)dE. \quad \dots\dots\dots\dots\dots\dots \quad (10)$$

This number will be made up of two groups; firstly, those electrons of energy E at t which suffer no radiation loss over dt (given by (9)); secondly, those electrons of energy greater than E at t which lose sufficient energy by collision and radiation to bring their energy down to $E-\beta dt$ at $t+dt$. Since this latter group of electrons loses energy βdt by collision in dt each must lose $W-E$ by radiation, if W is the energy of the electrons considered. The number of electrons at t with energy between W and $W+dW$ is $h(W,t)dW$. The number that lose energy $W-E$ by radiation in the distance dt is

$$h(W,t)\xi(W,W-E)d(W-E)dtdW. \quad \dots\dots\dots\dots \quad (11)$$

To obtain the total contribution to electrons in this group we must integrate over all values of W from E to $E_0-\beta t$ (the maximum possible energy at depth t). The number of electrons in the second group is thus

$$\int_E^{E_0-\beta t} h(W,t)\xi(W,W-E)d(W-E)dtdW. \quad \dots\dots \quad (12)$$

Equating (**10**) with the sum of (**9**) and (**12**) we get

$$h(E-\beta dt,t+dt)dE=h(E,t)dE-\int_0^E \xi(E,k)h(E,t)dEdtdk$$
$$+\int_E^{E_0-\beta t} h(W,t)\xi(W,W-E)d(W-E)dtdW. \quad \ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{13})$$

This equation may be reduced to the more convenient form

$$-\beta\frac{\partial h(E,t)}{\partial E}+\frac{\partial h(E,t)}{\partial t}=-h(E,t)\int_0^E \xi(E,k)dk+\int_E^\infty h(W,t)\xi(W,W-E)dW. \quad \ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{14})$$

It is permissible to extend the upper limit of integration in the last integral of (**13**) to ∞ since the function $h(E,t)$ is zero for all values of $E>E_0-\beta t$.

We now make use of the Mellin transform of the function $h(E,t)$, defined as

$$H(s,t)=\int_0^\infty E^{s-1}h(E,t)dE, \quad \ldots\ldots\ldots\ldots\ldots (\mathbf{15})$$

where s is a new variable. Multiplying equation (**14**) by E^{s-1} and integrating from 0 to ∞ of E, the left-hand side of (**14**) becomes

$$-\beta\int_0^\infty E^{s-1}\ \frac{\partial h(E,t)}{\partial E}dE+\int_0^\infty E^{s-1}\ \frac{\partial h(E,t)}{\partial t}dE$$
$$=-\beta\Big[E^{s-1}h(E,t)\Big]_0^\infty+\beta(s-1)\int_0^\infty E^{s-2}h(E,t)dE+\frac{\partial H(s,t)}{\partial t}$$
$$=\beta(s-1)H(s-1,t)+\frac{\partial H(s,t)}{\partial t}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{16})$$

The first term on the right-hand side of (**14**) becomes

$$-\int_0^\infty E^{s-1}h(E,t)dE\int_0^E \xi(E,k)dk. \quad \ldots\ldots\ldots\ldots (\mathbf{17})$$

We take for $\xi(E,k)$ the expression valid for complete screening:

$$\xi(E,k)dk=\left[1-\left(\frac{4}{3}+\alpha\right)\left(\frac{E}{k}-\frac{E^2}{k^2}\right)\right]\frac{k}{E^2}dk, \quad \ldots\ldots\ldots\ldots (\mathbf{18})$$

where $$\alpha=1/(9\log 183Z^{-1/3}). \quad \ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{19})$$

Expression (**17**) now becomes

$$-H(s,t)\int_0^1 \left\{\varepsilon-\left(\frac{4}{3}+\alpha\right)\left(1-\frac{1}{\varepsilon}\right)\right\}d\varepsilon, \quad \ldots\ldots (\mathbf{20})$$

with $$\varepsilon=k/E. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{21})$$

If the order of the integrations is changed, the second term on the right-hand side of (**14**) may be written:

$$\int_0^\infty E^{s-1}dE\int_E^\infty h(W,t)\xi(W,W-E)dW$$
$$=\int_0^\infty W^{s-1}h(W,t)dW\int_0^W \xi(W,W-E)\left(\frac{E}{W}\right)^{s-1}dE. \quad \ldots\ldots (\mathbf{22})$$

After substituting for $\xi(W,W-E)$ from (**4**) or (**18**) and making use of the fact that the terms inside the square brackets will depend only on the ratio $(W-E)/W$, we obtain for this term

$$H(s,t)\int_0^1 \left\{\varepsilon(1-\varepsilon)^{s-1}-\left(\frac{4}{3}+\alpha\right)\left(1-\frac{1}{\varepsilon}\right)(1-\varepsilon)^{s-1}\right\}d\varepsilon, \quad \ldots\ldots \quad (\mathbf{23})$$

with

$$\varepsilon=(W-E)/W. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{24})$$

Equations (**20**) and (**23**) are now combined to obtain for the right-hand side of the transformed equation

$$-H(s,t)A(s), \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{25})$$

where

$$A(s)=\int_0^1 \left\{\varepsilon-\left(\frac{4}{3}+\alpha\right)\left(1-\frac{1}{\varepsilon}\right)\right\}\left\{1-(1-\varepsilon)^{s-1}\right\}d\varepsilon. \quad \ldots \quad (\mathbf{26})$$

The complete equation for the transform $H(s,t)$ may now be written down. Combining (**16**) and (**25**) it is found to be

$$\partial H(s,t)/\partial t=-H(s,t)A(s)-\beta(s-1)H(s-1,t). \quad \ldots\ldots\ldots \quad (\mathbf{27})$$

If (**27**) is solved the original function $h(E,t)$ may be recovered by making use of the inversion theorem for the Mellin transform. The inversion of (**15**) is

$$h(E,t)=\frac{1}{2\pi i}\int_{\delta-i\infty}^{\delta+i\infty} E^{-s}H(s,t)ds, \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{28})$$

with the usual conditions on δ.

Examination of equation (**27**) shows that in the case $\beta=0$ we may write

$$H(s,t)=E_0^{s-1}e^{-A(s)t}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{29})$$

where the function has been normalized to one incident particle so that

$$h(E,0)=\delta(E_0-E)\,; \quad H(s,0)=E_0^{s-1}. \quad \ldots\ldots\ldots \quad (\mathbf{30})$$

Thus in the case $\beta=0$

$$h(E,t)=\frac{1}{2\pi iE_0}\int_{\delta-i\infty}^{\delta+i\infty}\left(\frac{E_0}{E}\right)^s e^{-A(s)t}\mathrm{d}s. \quad \ldots\ldots\ldots\ldots \quad (\mathbf{31})$$

In order to deal with cases for which β is not zero the solution (**29**) is generalized in the manner suggested by Bhabha and Chakrabarty (1943) to obtain

$$h(E,t)=\frac{1}{2\pi iE_0}\int_{\delta-i\infty}^{\delta+i\infty}\left(\frac{E_0}{E+\beta g(s,t)}\right)^s f(s,t,\beta)ds, \quad \ldots\ldots \quad (\mathbf{32})$$

where $g(s,t)$ and $f(s,t,\beta)$ are unknown functions and it is required that (**32**) shall go over into the form (**31**) for $\beta=0$.

Following the methods of Bhabha and Chakrabarty (1943) we obtain for the function $h(E,t)$

$$h(E,t)=\frac{1}{2\pi iE_0}\int_{\delta-i\infty}^{\delta+i\infty}\left\{\frac{E_0}{E+\beta[t-\frac{1}{2}\{A(s+1)-A(s)\}t^2+\frac{1}{2\cdot 3}\{A(s+1)-A(s)\}^2t^3]}\right\}^s$$

$$\left\{1-A(s)+\tfrac{1}{2}A^2(s)t^2-\frac{1}{2\cdot 3}A^3(s)t^3+\left(\frac{\beta}{E_0}\right)^2\frac{(s-1)(s-2)}{2\cdot 3}[2A(s-1)-A(s-2)-A(s)]t^3\right\}ds.$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{33})$$

If t is restricted to values less than unity, we may expand the denominator and neglect higher powers of t than the third. The resultant expression must now be evaluated after substitution of the value of $A(s)$ obtained from (**26**), viz.

$$A(s)=\left(\frac{4}{3}+\alpha\right)\left\{\frac{d}{ds}\log\Gamma(s)+\gamma-1+\frac{1}{s}\right\}+\tfrac{1}{2}-\frac{1}{s(s+1)}. \quad \ldots\ldots\ldots\ldots (34)$$

The contour integrals involved, although complicated, may all be evaluated using the standard methods in complex variable theory for the Bromwich contour. We obtain finally a very lengthy expression for $h(E,t)$. Numerical calculation showed that this expression could be shortened considerably, as many of the terms made almost negligible contributions to the desired function, but the simplified expression is still too long to quote here.

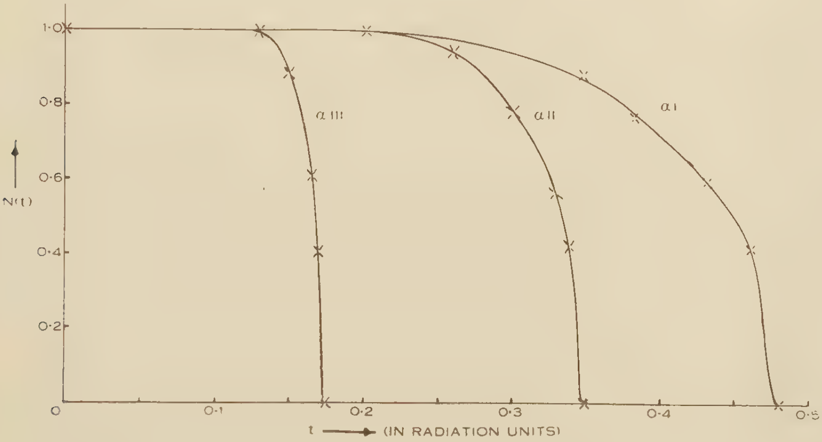


Fig. 1.—Integral range curves. The values of β/E_0 corresponding to the three curves are as follows: αI, 2·08; αII, 2·86; αIII, 5·73. (β=critical energy, E_0=initial energy.)

From this simplified expression it is possible to obtain numerical values of the function $h(E,t)$. For a given initial energy there exists an upper limit t_{max} on the penetration depth of the electrons. This limit is set by the definite amount of energy which the electrons lose by collision in reaching that depth. If we choose a number of values of t between 0 and t_{max}, the function $h(E,t)$ can be evaluated for each of these t values. In each case an appropriate range of values of E is selected. The distribution curve with respect to energy of the particles present at the particular depth is thus obtained for each value of t. This distribution, integrated over all values of E, gives the total number of particles $N(t)$ present at the depth t,

$$N(t)=\frac{1}{E_0}\int_0^{E_{max}} h(E,t)dE. \quad \ldots\ldots\ldots\ldots\ldots\ldots (35)$$

Here the factor $1/E_0$ is included to normalize to one incident particle and E_{max} is the maximum possible energy which a particle can have at depth t, given by

$$E_{max}=E_0-\beta t. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (36)$$

If this procedure is carried out for all the t values it is possible to draw a curve of $N(t)$ plotted against t. This represents the integral range curve, i.e. for a particular t value the number of particles with a range which exceeds this value. Curves for three values of the ratio β/E_0 are given in Figure 1. For water the three values of E_0 correspond to 20, 40, and 55 MeV. The differential range curves which correspond to the integral curves of Figure 1 are given in Figure 2.

III. Discussion of Results

A comparison of our results with those of Corben (1938) or Eyges (1949, 1950) for higher values of initial energy is difficult because these authors give curves showing the dependence of the integral straggling function on the penetration depth, i.e. from their curves one could obtain for a given value of t the integral of $h(E,t)$ between the limits E and E_{max}, but only for values of E not too far below the initial energy E_0. In using this quantity to obtain integral and differential range curves as described above, one would have to neglect the contribution of all electrons with energies below the value E, and for the results given by both authors the error introduced in this way would be both large and indeterminate. From the work of Eyges we have drawn approximate curves for initial energies slightly above the critical energy. As might be expected, the differential range curve is much broader than those of Figure 2, rising in almost a straight line from zero to the maximum and having a half

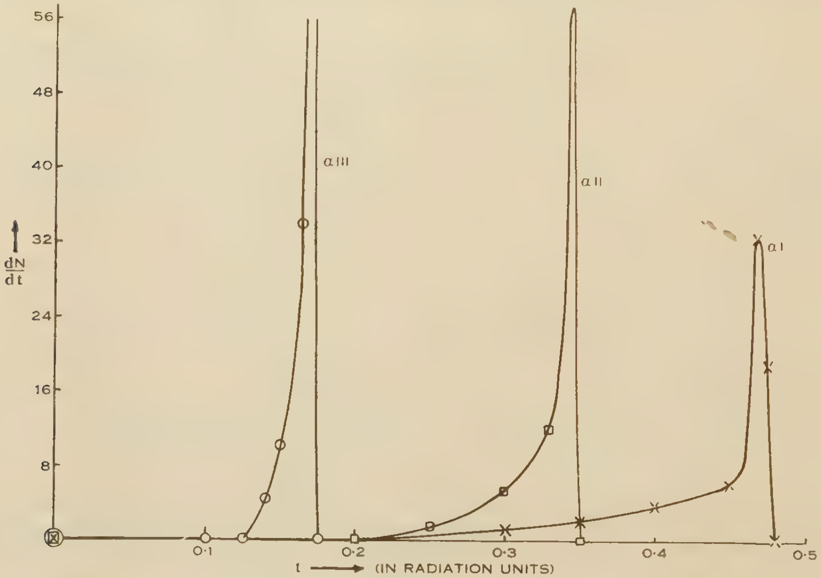


Fig. 2.—Differential range curves corresponding to the integral curves of Figure 1. The values of β/E_0 corresponding to the three curves are as follows: αI, 2·08; αII, 2·86; αIII, 5·73. (β=critical energy, E_0=initial energy.) The maximum of αIII occurs at $dN/dt=102$.

width of about 0·35 radiation units as compared with a half width of only 0·0125 for the curve αI of Figure 2.

The extent of straggling due to radiation energy loss is shown in Figure 2. Most of the electrons possess a range which is not very different from the most probable value, and only a relatively small number have ranges which are appreciably below this value. It might seem at first sight, therefore, that this straggling would be unimportant for practical purposes, for example with regard to its effect on the range-energy relation below the critical energy. A convenient method of measuring the energy of a beam of fast electrons is by means of absorption measurements, making use of the range-energy relation, but if the electron beam is heterogeneous in energy, the distortion of the range spectrum by radiation straggling may be serious.

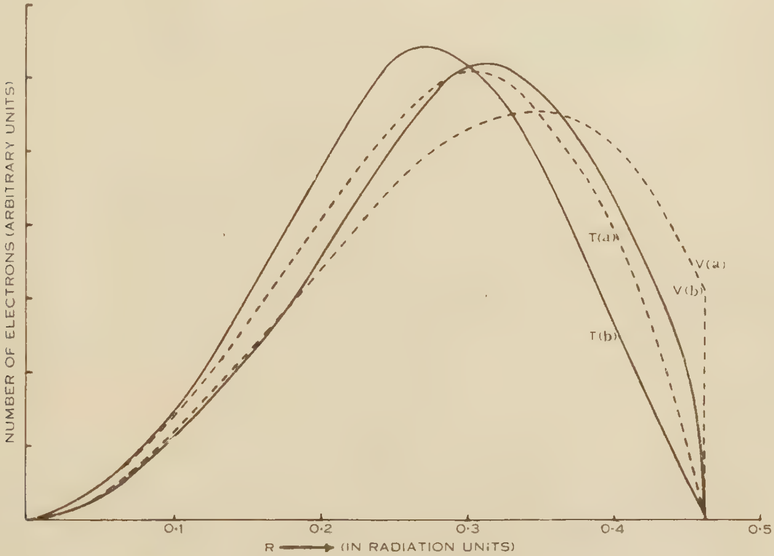


Fig. 3.—Curves showing the effect of radiation straggling on two of the theoretical curves of Tiomno and Wheeler. $V(a)$ and $T(a)$ are respectively the vector coupling theory curve and the tensor theory curve without correction; they represent number of electrons plotted against range. $V(b)$ and $T(b)$ are the curves after correction for the effects of radiation straggling.

An important case would be the determination of the energy spectrum of meson decay electrons by absorption measurements in layers of light material of different thicknesses. A knowledge of the precise form of the spectrum is required in order to distinguish between various meson field theories which have been applied to the decay of the μ-meson, for these theories lead to broad energy distributions for the decay electrons which do not differ very greatly from one another. Straggling may broaden the curves further and increase the similarity so much as to make it extremely difficult to distinguish between them.

To illustrate this, we have taken two representative theoretical energy distributions for meson decay electrons from the work of Tiomno and Wheeler

(1949). One curve ($V(a)$ in Fig. 3) has been calculated on the basis of the vector coupling theory and the other (curve $T(a)$) on the tensor coupling theory. In the former the abrupt vertical drop at the high energy end of the energy distribution is a feature providing the most sensitive means of differentiating it from the other distribution. It should be noted that in Figure 3 the abscissae have been adapted to correspond to range instead of energy. The two curves have been modified by applying the correction due to radiation straggling and the resulting curves, which would be obtained from measurements made with absorbing screens if straggling due to scattering were negligible, are also shown in Figure 3 ($V(b)$ and $T(b)$ respectively). It is obvious that the two corrected curves resemble one another much more closely than do the two original distributions, the discontinuity in the vector case having been ironed out and the difference in slope of the curves having been almost eliminated. The experimental difficulty which will arise in attempting to discriminate between such curves is obvious.

In an interpretation of experimental results, or in the design of experiments the straggling due to scattering must also be taken into account. From results quoted by Steinberger (1949) it would appear that, for electrons with energies between about 40 and 55 MeV. in water, the straggling due to scattering is of the same order of magnitude as the radiation straggling. For electrons with initial energies of the order of 20 MeV. in water the radiation straggling can be seen to be quite small, whereas the straggling caused by the scattering should become more important for electrons of lower energy.

IV. ACKNOWLEDGMENTS

The author wishes to thank Associate Professor C. B. O. Mohr for many helpful discussions during the course of this work. He is also indebted to Mr. V. Officer and Mr. G. Mauldon for advice concerning the experimental implications of some of the results obtained.

V. REFERENCES

BHABHA, H. J., and CHAKRABARTY, S. K. (1943).—*Proc. Roy. Soc.* A **181** : 267.
CORBEN, H. C. (1938).—*Proc. Camb. Phil. Soc.* **34** : 540.
EYGES, L. (1949).—*Phys. Rev.* **76** : 264.
EYGES, L. (1950).—*Phys. Rev.* **77** : 81.
STEINBERGER, J. (1949).—*Phys. Rev.* **75** : 1136.
TIOMNO, J., and WHEELER, J. A. (1949).—*Rev. Mod. Phys.* **21** : 144.

OBSERVATIONS OF GALACTIC RADIATION AT FREQUENCIES OF 1210 AND 3000 Mc/s.

By J. H. Piddington* and H. C. Minnett*

[*Manuscript received May* 23, 1951]

Summary

Measurements are described of galactic radiation at frequencies of 1210 and 3000 Mc/s. The results provide extensions of previous radio emission spectra by more than $2\frac{1}{2}$ octaves.

The intensity of radiation from near the galactic centre was measured and an estimate made of the direction of the maximum of such radiation. A new " discrete source " of peculiar spectrum was discovered very close to the centre of the Galaxy. Evidence suggests that the power output of this and some other sources in the radio spectrum may exceed the *total* power output of the Sun.

Radiation was also observed from the direction of the Crab Nebula, from the known " radio source " in Centaurus, and from the Moon. Two nebulae (M 31 and NGC 7293) were investigated with negative results. The spectra of two sources are given between the limits 18·3 and 1200 Mc/s.

I. Introduction

Observations of galactic radio-frequency radiation have been made by a number of observers using frequencies between 9·5 Mc/s. (Friis and Feldman 1937) and 480 Mc/s. (Reber 1948). In addition to the general background of radiation, numerous " discrete sources " have been observed at frequencies up to 160 Mc/s. (Stanley and Slee 1950). The available data have allowed spectrum curves to be constructed for the general background radiation (Piddington, in press) and the discrete sources (Stanley and Slee 1950), over the frequency ranges indicated. Extension of these spectra is desirable as this may provide evidence about the source of the radiation. Some of the data given below have already been used by Piddington (in press) in this connection. It was with this objective that measurements described below were undertaken at frequencies of 1210 and 3000 Mc/s. Unfortunately the experimental difficulties involved increase rapidly with frequency, and, as will be seen below, were considerable. However, the results provide appreciable extensions of the spectra measured. In some cases the results obtained were negative, in that the flux density of radiation was below the observable limit. These have some value in providing upper limits.

II. Experimental Methods

The receiving systems used on both 1210 and 3000 Mc/s. employ the synchronous switching principle described by Dicke (1946) and embodied in the

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

24,000 Mc/s. radiometer of Piddington and Minnett (1949), and in subsequent equipments for 10,000 and 3000 Mc/s. Certain new features and techniques developed for the present investigations are discussed below.

(*a*) 1210 *Mc/s. Equipment*

In the radiometers previously used by the authors, the radiation received by the aerial was compared 25 times per second with the radiation from the dummy load. Output drifts due to variations of receiver gain are therefore proportional to the difference between the received power and that emitted by the dummy load. If this difference is small, the output drifts may be much smaller than those in a direct coupled equipment which are proportional approximately to the total receiver noise power. A switched receiving system which automatically maintains the difference equal to zero has been developed by Ryle and Vonberg (1948) for solar measurements. In the present investigations, the power received by the aerial was very small and was composed of radiation from the source being investigated, leakage into the aerial from terrestrial surroundings, and emission by the feeder system. A second aerial, directed at a " cold " and uniform part of the sky, was therefore used as the dummy load. It was then possible to arrange for the two aerial temperatures to differ only when the source under investigation was in the beam of the " main " aerial. Two different aerials have been used as the main aerial. The first was 10 ft. in diameter ; the second, which had an aperture of 18 by 16 ft., has been described by Lehany and Yabsley (1949). As a reference aerial, a dipole-fed parabolic reflector of aperture 5·5 ft. was used.

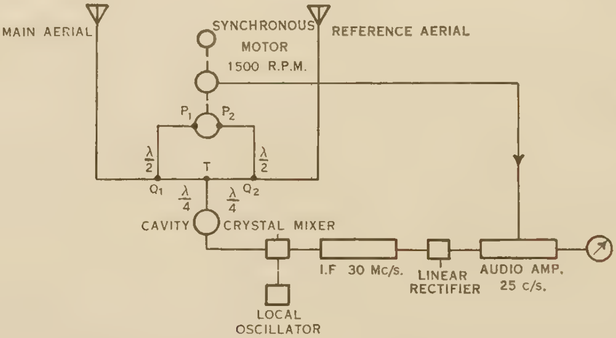


Fig. 1.—Block diagram of the 1210 Mc/s. receiving system.

The rotating-disk waveguide switch used in the higher frequency radiometers was replaced by a compact, coaxial-line, rotating-condenser switch of the type often used in metre-wave radar direction finding. This switch is rotated at 1500 r.p.m. by a synchronous motor, which also drives a permanent-magnet generator delivering 50 volts A.C. to the synchronous rectifier of the receiver (Fig. 1). The switch terminals P_1 and P_2 are connected by half-wave lines to Q_1 and Q_2. The operation is such that a very low impedance appears across Q_1 and a very high impedance across Q_2 for 135° of each switch revolution. One half-revolution later these impedance conditions are interchanged and the cycle of operations is repeated 25 times per second. Since Q_1T and Q_2T are quarter-wavelengths, the receiver at T is connected in turn to the main aerial and reference aerial. Each aerial matches the characteristic impedance of its feeder so that the receiver is presented with this impedance except during the switch transitions.

The switch itself is shown in more detail in Figure 2. The semi-circular rotor meshes in turn with the stator plates C_1 and C_2. The shape of the plates ensures that the transition periods between the two conditions are as small as possible (45° each). In the position shown, the capacity at C_1 is a maximum and the length of coaxial line C_1P_1 is chosen to produce a near short circuit at P_1.

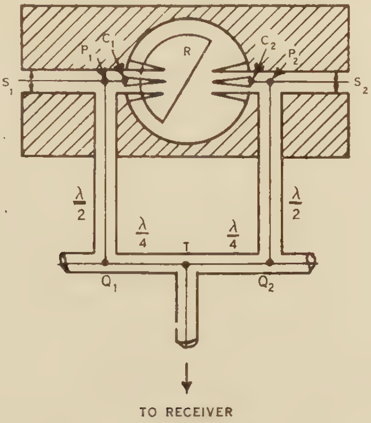


Fig. 2.—The rotating condenser switch.

At the same time, the capacity at C_2 is a minimum and the plunger S_2 is tuned to resonate the line $C_2P_2S_2$. The impedance at P_2 is therefore very high. When the rotor is meshed with C_2, these impedance conditions are interchanged, the adjustments now being the location of P_2 and the tuning of S_1. The two sets of adjustments are substantially independent.

Figure 1 shows a resonant cavity connected between the switch and the crystal mixer. This has a loaded Q of about 100 and is tuned to the signal frequency of 1210 Mc/s. Its purpose is to eliminate the image and harmonic responses of the receiver in order to avoid calibration errors of the type discussed by Lehany and Yabsley (1949). The cavity also discriminates against leakage of local oscillator power into the rotating switch circuits, with consequent modulation of the crystal current and the production of a troublesome output signal.

Stability of the frequency of the local oscillator is important in avoiding drifts when the two aerials have different frequency characteristics. Satisfactory results were achieved in practice by maintaining the oscillator at a fairly constant temperature; the high tension and heater voltages were obtained from the electronically regulated units supplying the whole equipment.

The noise factor of the receiver is approximately 11 decibels and the intermediate frequency bandwidth is 7 Mc/s. A maximum current gain of 2000 is available between the linear second detector and the recording output meter. The effective time constant of this meter may be varied between 0·5 and 72 seconds.

The theoretical r.m.s. output fluctuation for the longest time constant is 0·1 °K., and with the two aerials replaced by well-matched loads maintained at the same temperature, this limit is readily achieved. During a 24-hour run, long-period drifts do not occur faster than 0·1 °K./hour. The minimum signal which may be detected in practice depends on the type of measurement and the averaging process used. In the observations described in this paper, a temperature change of about 0·4 °K. was detectable.

With the two aerials in operation, drifts as fast as 5–10 °K./hour have been observed at times of rapid ambient temperature change, e.g. at sunrise and sunset. It is thought that most of this is due to differential temperature variations of the two feeder systems. Other variations were ascribed to differential leakage of solar radiation into the aerial systems. Observations were therefore confined as much as possible to the night hours.

The sensitivity of the equipment is found by replacing both aerials by matched loads. One of these is a coaxial-line carbon load which may be heated through a calibrated temperature interval (about 50 °K.). For rapid monitoring of the sensitivity and adjustment of the equipment, a hot-wire lamp operating at approximately 1600 °K. has also been used (Lehany and Yabsley 1949).

The power gain of the 10-ft. aerial relative to an isotropic radiator is 625. This result was found by comparison with a smaller parabola which had been calibrated against a dipole having an assumed gain of 1·67. A cross-check was provided by measuring the polar pattern of the feed to the 10-ft. aerial. The gain was then calculated assuming uniform phase distribution over the aperture and allowing for the proportion of the radiated energy missing the reflector; the result was 16 per cent. higher than that found by direct measurement. The proportion of the total energy accepted by the aerial and contained within the main beam is 60 per cent. and the beam width to the half-power points ±2·9°.

The power gain of the 18-ft. by 16-ft. aerial relative to an isotropic radiator is 2400. This figure followed from a comparison with the 5·5-ft. aerial, the gain of which had been calibrated by transmission and comparison tests with a 30-in. parabola. The beam width of the 18-ft. by 16-ft. aerial is ±1·4° to half-power, and 65 per cent. of the total accepted energy is calculated to be within the main beam and near side lobes.

(*b*) 3000 *Mc/s. Equipment*

This radiometer has been described by Piddington and Hindman (1949) in connection with solar observations. For the present measurements, the aerial aperture was increased to 7·5 ft. by fitting extensions to the existing parabola. With this modification the characteristics of the aerial were: power gain over an isotropic radiator 1500; beam width to half-power ±1·7°; and proportion of energy in the main lobe 80 per cent. The gain was measured by transmission and comparison tests using two horns. The remainder of the equipment was unchanged.

(c) *Observational Techniques*

At the frequencies concerned, the flux density of radiation to be measured is very low and the aerial temperature change is usually only a few degrees Kelvin. Two different observational techniques were used. In the direct method, the aerial beam was set ahead of the source, which was then allowed to drift through it. The longest possible time constant was used consistent with the faithful recording of the passage of the source through the beam. This method was used only at 1210 Mc/s. The 18-ft. by 16-ft. aerial and a 72-second time constant allowed a minimum flux density of 10^{-24} W. m.$^{-2}$ (c/s.)$^{-1}$ to be detected after allowing for feeder losses. Figure 3 is a record of the passage of the aerial beam through part of the central region of the Galaxy (Dec. 30 °S.). The sharp central maximum is due to radiation from a new discrete source discussed in Section III. The spread maximum corresponds to the general background of galactic radiation. The random fluctuations are instrumental and have been discussed in Section II. The maximum increase of equivalent temperature at the receiver input was 11 °K. Also shown in Figure 3 is the aerial beam width between half-power points (corrected for the Declination of 29°).

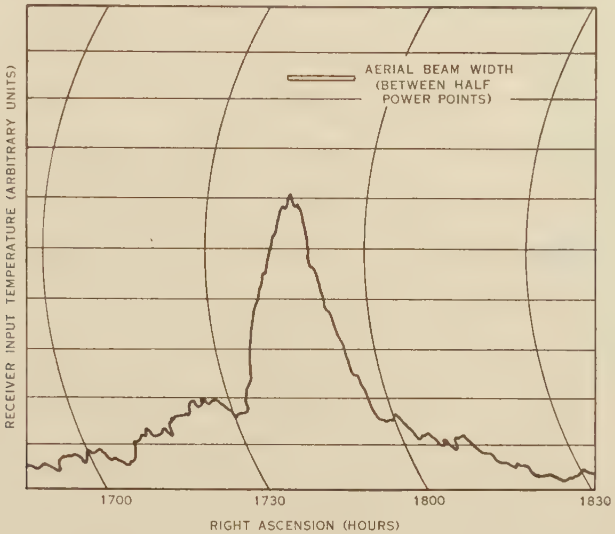


Fig. 3.—Record of the galactic plane and the Sagittarius-Scorpius source passing through the 1210 Mc/s. (±1·7°) aerial beam.

A second technique, called the " beam-swinging " method, was used for a few preliminary observations at 1210 Mc/s. and for all of the 3000 Mc/s. observations. With a receiver time constant of about 10 seconds, the aerial beam was pointed in a given direction relative to the Earth (position *A*). After an interval of one or two minutes, the angle of the beam was changed to a position some degrees west of position A (position B) and left for the same interval of time before being returned to position A. This procedure was repeated for the time taken for the source to move from position A to position B. The difference in aerial temperature between the two positions was measured on each occasion and plotted against time. If the presence of the source in the aerial beam causes an increase in aerial temperature of T_A, and if the background of radiation is assumed uniform, then the curve of aerial temperature difference varies through a total range of $2T_A$ as the source moves from A to B. It is estimated that using this method the minimum flux density detectable with the 3000 Mc/s. equipment under ideal conditions was 10^{-23} W. m.$^{-2}$ (c/s.)$^{-1}$. The method was not used with the large aerial at 1210 Mc/s. because movement of the aerial caused random fluctuations of apparent aerial temperature.

(*d*) *Reduction of Results*

In discussing radiation from galactic sources which are uniformly " bright ", the source may be described in terms of the brightness B (W. m.$^{-2}$ (c/s.)$^{-1}$ sterad.$^{-1}$), or the brightness temperature T_b of the region. The relation between the two units is given by $B=2kT_b/\lambda^2$, where λ is the wavelength. When an aerial is directed towards a region of uniform brightness over the (main) beam and zero brightness elsewhere, the aerial temperature T_a is equal to $T_b\eta$ where η is the proportion of the total radiated power lying within the beam when the aerial is used as a radiator. If the region is not uniformly bright over the beam, T_a/η is equal to the weighted mean brightness over the beam.

The flux density S of radiation at the aerial from sources subtending a solid angle small compared with that of the aerial beam, is related to T_a by the equation

$$S=\frac{8\pi}{G}\,\frac{kT_a}{\lambda^2},$$

where G is the power gain of the aerial compared with an isotropic radiator.

III. Radiation from Near the Galactic Centre

At the frequencies concerned it is at present impossible to measure the very low absolute temperatures of the " cold " parts of the Galaxy. The best that can be provided is a temperature difference between two parts, one of which may be a region of particular interest such as the centre of the Galaxy and the other a " cold " part. Since radiation from the latter is only a small proportion of that from the former (Piddington, in press), the method provides a reasonable estimate of the brightness temperatures of regions of maximum radiation.

Contour charts of regions of galactic radiation by Reber (1944, 1948), Bolton and Westfold (1950*a*), Allen and Gum (1950), and others indicate that the maximum of radiation comes from near the direction R.A. 17 hr. 50 min., Dec. 27°S. ($l=330°$, $b=-2°$). It was decided to measure the excess radiation at 1210 and 3000 Mc/s. with the aerials pointed in that direction.

(*a*) *Observations at* 1210 *Mc/s.*

Preliminary observations of radiation from near the galactic centre, using the 10-ft. diameter aerial (beam width ±2·9°), were made in 1948. The receiver sensitivity was sufficiently low to warrant the use of the beam-swinging method of observation described in Subsection II (*c*). The beam was set at Declination 27 °S. and alternated between two positions 72 minutes apart in hour angle, while the plane of the Galaxy passed between these positions.

After correcting for side lobes, the excess (average) brightness temperature at the galactic centre over the mean for the other directions was 15·3 °K. The result was the mean of two, which differed by 8 per cent. An allowance of 1·1 °K. must be made for the discrete source at Declination 30 °S. (see below), reducing the value to 14 °K. The main uncertainty was in determining aerial characteristics. The overall uncertainty is probably about 20 per cent.

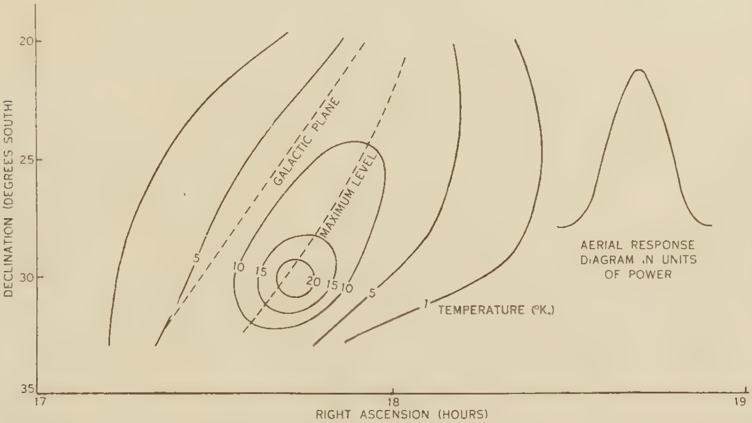


Fig. 4.—Aerial temperatures at 1210 Mc/s. in the region of the galactic centre.

During 1950 the large aerial (beam width ±1·4°) became available during restricted periods for galactic observations. It was decided to conduct a limited survey of the region in the vicinity of the galactic centre. For a number of declinations (spaced about 1° apart) and with the aerial beam well above the horizon, the region concerned was allowed to pass through the beam, and aerial temperature was recorded continuously. The results were then reduced to a set of equal intensity contours shown in Figure 4.

An examination of these contours reveals several points of interest. Firstly, the sharp maximum at R.A. 17 hr. 44 min., Dec. 30 °S. appears to be a new, and remarkably powerful, discrete source. It is discussed more fully below. Secondly, the contours provide another determination of the position of the galactic centre. This is also discussed below. Finally, the results allow a determination to be made of the excess radiation from near the galactic centre (above an unknown, but probably low, level away from the galactic plane).

Subtracting the effect of the discrete source, the maximum equivalent aerial temperature was 11 °K. The value of η for the aerial was 0·65 (Section II) so that the weighted mean brightness temperature of the region was 17 °K. and the brightness $7 \cdot 6 \times 10^{-21}$ W. m.$^{-2}$ (c/s.)$^{-1}$ sterad.$^{-1}$. The uncertainty is probably about 20 per cent. The result is in satisfactory agreement with that obtained with the smaller aerial and different technique. The temperature difference measured with the larger aerial should be larger than that for the smaller aerial because its narrower beam would be concentrated more effectively on the maximum of intensity.

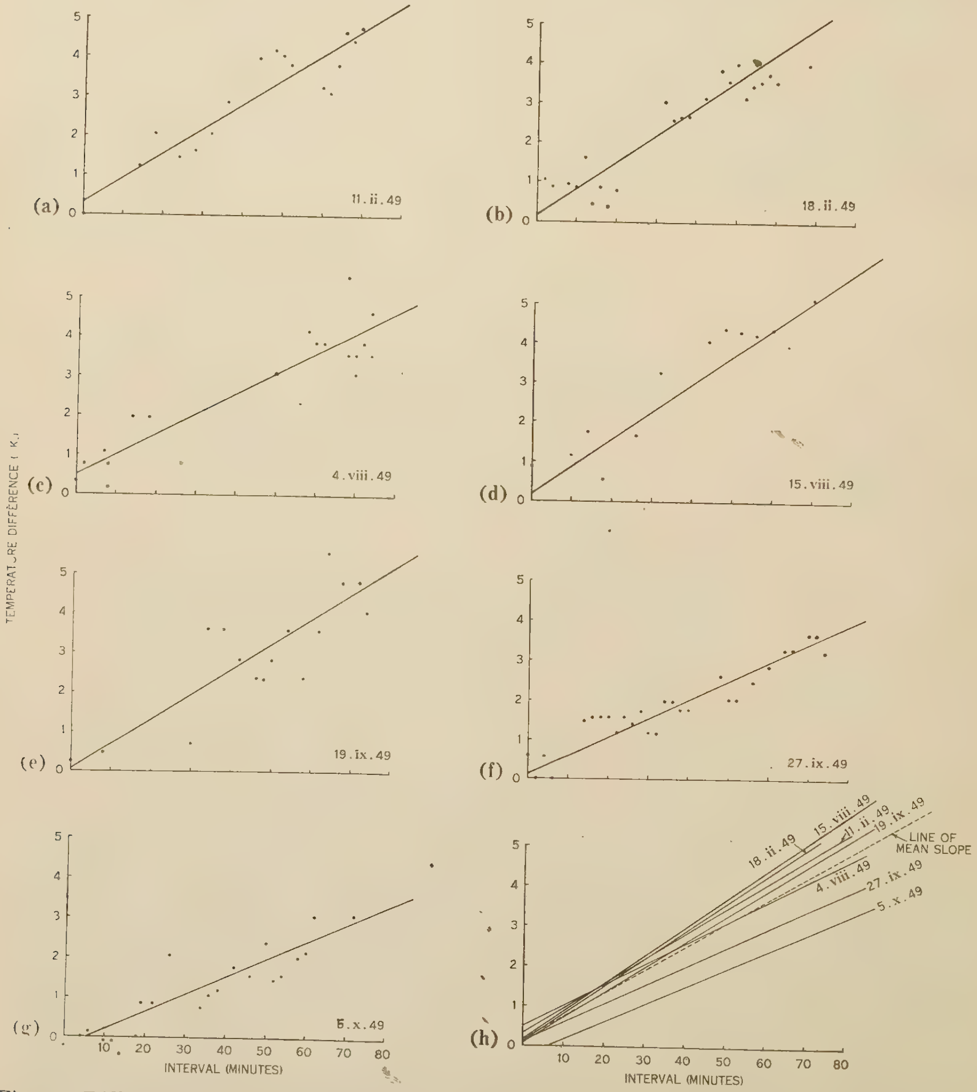


Fig. 5.—Difference in aerial temperatures at 3000 Mc/s. for two beam positions 18° apart (both at Declination 27 °S. as the galactic plane crosses the beams).

To find the actual brightness temperature, it would be necessary to make an allowance for the variation of aerial response within the main beam, thereby slightly increasing the maximum. In view of the narrowness of the beam this hardly seems warranted, but if such allowance were required it could be made in the manner of Hey, Parsons, and Phillips (1948) or Bolton and Westfold (1950*a*).

(*b*) *Observations at* 3000 *Mc/s.*

Owing to the extremely weak radiation to be measured, it was only possible to use the beam-swinging method at this frequency. The aerial (beam width ±1·7°) was pointed in the same directions as for the 1210 Mc/s. observations mentioned above. Numerous sets of observations were made, more than half being rejected because of excessive zero drift or other signs of instability.

For each run a series of temperature differences was obtained and plotted against the time elapsed from the beginning of the experiment. A straight line was therefore drawn as a best fit through each set of points. In the absence of any radiation, the line should be parallel to the time scale, but a radiating source crossing the beam would result in changes in the temperature differences and a finite slope of the line. The overall change in temperature difference provides a measure of the aerial temperature caused by the source.

The results of seven runs were used and are shown in Figure 5, graphs (*a*) to (*g*). The result of each experiment depends only on the slope of the resulting line, not on its level. All the lines are plotted in Figure 5 (*h*) together with a dashed line corresponding to the mean slope. The scatter of slopes is seen to be reasonably small. The required numerical result is given by half the temperature difference for points on the line of mean slope separated by a time interval of 72 minutes ; it is 2·1 °K. The factor η had a value 0·80, so that the excess brightness temperature near the galactic centre and for the beam concerned was 2·6 °K., or, in terms of source brightness, 7×10^{-21} W. m.$^{-2}$ (c/s.)$^{-1}$ sterad.$^{-1}$.

The accuracy of each determination of temperature difference is probably low as indicated by the scatter of results. However, as a number of runs were made it is likely that, as for the 1210 Mc/s. experiment, the overall accuracy is limited by the uncertainty in the aerial measurements to about 20 per cent. Owing to the narrow aerial beam the effect of the discrete source at Declination 30 °S. should be negligible.

(*c*) *The Discrete Source in Sagittarius-Scorpius**

The position of the discrete source is R.A. 17 hr. 44 min., Dec. 30 °S. ($l=328°$, $b=-3°$) with an estimated uncertainty in Right Ascension of about 2 minutes and in Declination of about 1°.

Only an upper limit of the size of the source can be derived from the data. The passage of a source of finite size through the aerial beam produces a record which is broader than for a point source. It may be shown, however, that the effect is negligible when using the ±1·4 beam for sources smaller than 1·5° in diameter. No widening effect could be detected for the Sagittarius-Scorpius source, so that its diameter has an upper limit of 1·5°.

---

* The best determination of the position is actually on the border between the constellations of Sagittarius and Scorpius.

The excess aerial temperature due to the source was 11·3 °K. so that the power flux density was $2 \cdot 6 \times 10^{-23}$ W. m.$^{-2}$ (c/s.)$^{-1}$ with an uncertainty of about 20 per cent. Although this flux density indicates a particularly powerful source (at 1210 Mc/s.), its presence has not previously been reported at lower frequencies. Reber's (1948) failure to observe it at 480 Mc/s. may have been due to the fact that it is situated near the extreme southerly limit of his survey.

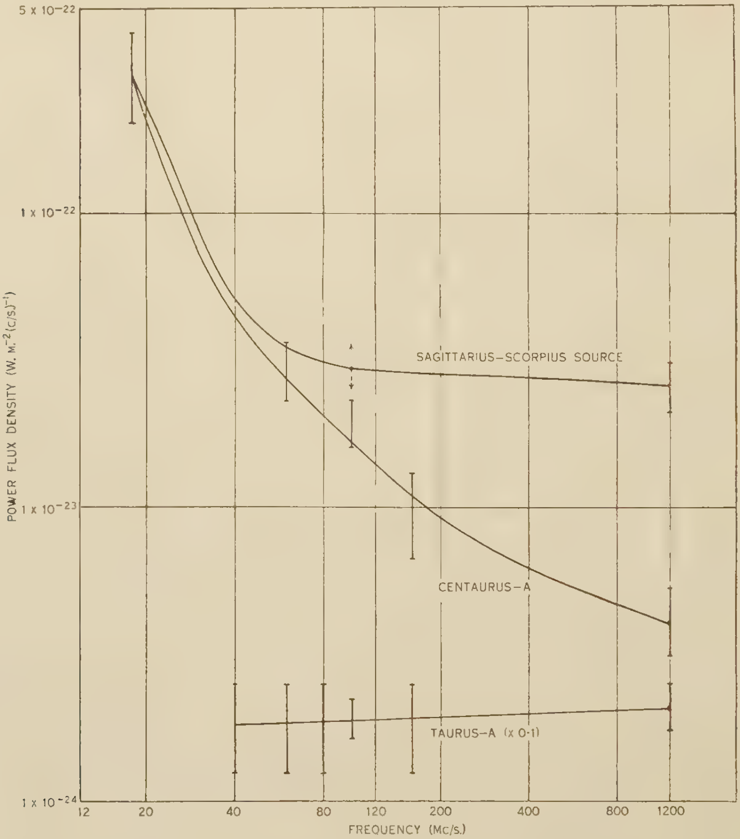


Fig. 6.—Spectrum of radiation from the discrete source in Sagittarius-Scorpius, from Taurus-A, and from Centaurus-A. The flux-density scale for Taurus-A has been changed by a factor of 10 to separate the curves.

Since the source was first investigated, B. Y. Mills of this Laboratory has provided preliminary information (personal communication) of a source at about R.A. 17 hr. 50 min., Dec. $28\frac{1}{2}$ °S. observed with an interferometer at 100 Mc/s. The flux density was, very approximately, 3×10^{-23} W. m.$^{-2}$ (c/s.)$^{-1}$ the uncertainty being large and indefinite. In spite of the difference of $1\frac{1}{2}$° in the apparent position, it seems likely that this is the same source. C. A. Shain of this Laboratory has also observed what appears to be the same source at 18·3 Mc/s., finding a flux density of about 3×10^{-22} W. m.$^{-2}$ (c/s.)$^{-1}$ with an unknown uncertainty (probably not worse than 50 per cent.). The spectrum of radiation, determined by the three measurements, is shown in Figure 6.

The spectrum of the Sagittarius-Scorpius source is not yet sufficiently definitely defined to warrant a detailed discussion. It is, however, worth pointing out that between the limits of 100 and 1200 Mc/s. the spectrum resembles that of an optically thin, thermally radiating gas (and also the Taurus-A spectrum as indicated below). It is evidently desirable to verify the low-frequency end of the spectrum.

An interesting feature of the Sagittarius-Scorpius source is that it lies close to the galactic plane and very close (within experimental uncertainty) to the plane defined by the maximum level of radio emission. It also lies very close to the centre of the Galaxy and to the maximum of galactic radiation; the latter is discussed in the following subsection. The significance of the position may be considerable. If the source were relatively close to the Sun it could lie in any direction with equal probability. If, on the other hand, it were at a distance from the Sun which was a considerable fraction of the galactic diameter, it would be more likely to lie in a direction close to the plane of the Galaxy.

The two strongest sources yet located, one in Cassiopoeia (Ryle and Smith 1948) and one in Cygnus (Stanley and Slee 1950) also lie close to the galactic plane: at 3 and 5° respectively. An examination of the results of Stanley and Slee shows several other sources lying close to the galactic plane, a much greater proportion than would be expected if they were uniformly distributed. The implication is that some sources lie at distances much greater than the thickness of the Galaxy in the vicinity of the Sun, distances of the order of 10^4 parsecs.

If sources do lie at these great distances, it follows that they must emit an enormous amount of energy in the radio spectrum. Thus, the Sagittarius-Scorpius source, if distant 10^4 parsecs, must emit 4×10^{19} W. (c/s.)$^{-1}$. In the whole radio spectrum (say 1000 Mc/s., although it may be considerably more) the power radiated is 4×10^{28} W., about 100 times greater than the *total* power radiated by the Sun.

Although, as seen above, the accuracy of location of the source is not high, it may be significant that the position found almost coincides with that of the galactic nebula NGC 6451 (H VI 13), a loose cluster of about 70 stars extending over 15 minutes of arc. A much more accurate determination of position is required, however, before the coincidence is given serious consideration.

(*d*) *The Position of the Maximum of Galactic Radiation*

After subtracting the effect of the small intense source discussed above, the best estimate of the position of the maximum of radiation is at R.A. 17 hr. 50 min., Dec. 27½ °S. ($l=330°$, $b=-2\frac{1}{2}°$). Systematic errors should have been largely eliminated since the aerial beam was lined up on the Sun before each run or before each series of two to three runs. The uncertainty in the direction of changing right ascension may be as low as ½° but that in the other direction is much higher, perhaps as much as 3°. The reasons for these different degrees of accuracy lie in the method of measurement, which allows a comparison of intensities at different declinations to be made only by comparing the results of different sets of observations between which changes in receiver gain may have occurred. On the other hand, a series of measurements at different right ascensions is made during a single run.

Positions of the maximum intensity of radiation at other frequencies deduced from the published contour charts are listed in Table 1. Not all published data are included because a useful determination is not possible in some cases.

TABLE 1

DETERMINATIONS OF THE DIRECTION OF THE MAXIMUM OF GALACTIC RADIATION

Observer	Frequency (Mc/s.)	Position			
		R.A.	Dec.	l	b
Bolton and Westfold (1950*a*)	100	17 hr. 33 min.	—30°	326°	—1°
Reber (1944)	160	17 50	—25°	332°	—2°
Allen and Gum (1950) ..	200	17 30	—35°	322°	—3°
Reber (1948)	480	17 53	—24°	333°	—2°
Present paper	1210	17 50	—27·5°	330°	—2·7°

It is evident from Table 1 that there is a scatter of more than 11° in the estimated positions. This is almost entirely in the values of galactic longitude, the values of galactic latitude falling within ±1°. It seems probable that the scatter is mainly due to experimental errors, being larger in l than in b because of the method of making the measurements described above.*

Northcott and Williamson (1948) and Bolton and Westfold (1950*b*) have attempted to determine features of galactic structure from radio data. However, Piddington (in press) has shown that the radio contours, even when allowance is made for possible experimental errors, change quite markedly with frequency. He also showed that the radiation probably has at least two different mechanisms of origin which give entirely different spectra. This means that galactic structure, determined by the method of Bolton and Westfold, *must* depend on the radio frequency used. It also means that the position of the galactic centre as determined above *may* depend on the radio frequency.

To settle this last point a number of accurate surveys, from which systematic errors have been eliminated, are needed. Narrow aerial beams should be used to facilitate interpretation of the experimental data. As well as the main survey on each frequency, an interferometer search for discrete sources should be made. Since these may have peculiar spectra and finite size they constitute a considerable danger to interpretation of observations with wide-beam aerials.

---

* In a paper which has appeared since the present one was written, Seeger and Williamson (1951) have published the results of careful 205 Mc/s. observations. In the region of galactic longitude 330° their maximum of radio emission has an (average) value of —1·6°. The value computed from their determination of the position of the galactic pole from the radio data is —2·4°. Both are in satisfactory agreement with the data of Table 1.

IV. RADIATION FROM THE CONSTELLATION OF TAURUS

Diffuse radio-frequency radiation from the Milky Way in Taurus has been observed by Reber (1944, 1948) and others who have prepared contour charts of levels of galactic radiation. More recently, a source of radiation identified with the Crab Nebula (NGC 1952) has been found by Bolton and Stanley (1949) and designated Taurus-A. Its spectrum in the interval 40 to 160 Mc/s. has been determined by Stanley and Slee (1950). This spectrum, unlike that of the other discrete sources hitherto investigated, does not show any change in the flux density of radiation received as the frequency is varied. Such a spectrum curve, as Townes (1947) has shown, is typical of radiation from an optically thin cloud of ionized gas. It is of interest, therefore, to extend the spectrum of the Crab Nebula to as high a radio frequency as possible and see whether the flux density of radiation still remains constant.

(*a*) *Observational Results*

On May 11, 1950, observations were made at a frequency of 1210 Mc/s. of radiation from the direction of the Crab Nebula. The large aerial (beam width $\pm 1\cdot 4^\circ$) was used, and the Sun was used to line up the aerial beam. The beam-swinging technique was employed with a displacement between the beam positions of 4°, which was sufficient to reduce the aerial response to a negligible value. The results of three runs had a spread of $8\cdot 2$ per cent. and the mean value of flux density deduced was $2\cdot 1\times 10^{-23}$ W. m.$^{-2}$ (c/s.)$^{-1}$. The uncertainty in this result is probably about 20 per cent.

On several occasions attempts were made to observe and measure radiation from Taurus-A at 3000 Mc/s. The results were not conclusive, although some signs of radiation seemed to be present. The upper limit of flux density was estimated as about 2×10^{-23} W. m.$^{-2}$ (c/s.)$^{-1}$.

(*b*) *Discussion*

The spectrum of Taurus-A (with the flux density reduced by a factor of 10 so as to separate the curve from the other spectra) is shown in Figure 6. The data for 40 to 160 Mc/s. have been given by Stanley and Slee (1950). The curve, as far as 1210 Mc/s., remains a nearly horizontal straight line. Such a spectrum is consistent with a thermal origin of the radiation in an optically thin layer of ionized gas and it is of interest to examine this possibility.

The brightness temperature of the Crab Nebula corresponding to a flux density S of $1\cdot 85\times 10^{-23}$ W. m.$^{-2}$ (c/s.)$^{-1}$ at 40 Mc/s. is given by

$$T_b = S\lambda^2/2k\Omega,$$

where λ, the wavelength, is $7\cdot 5$ m., k is Boltzmann's constant, and Ω the solid angle subtended by the source. Bolton and Stanley (1949) have shown that the size of the radio source (in one dimension only) is not greater than that of the optical source. The value of Ω is found by assuming that the dimensions are 6 min. arc by 4 min. arc (Baade 1942), so that $\Omega = 1\cdot 6\times 10^{-6}$ sterad. and $T_b = 2\cdot 4\times 10^{7}$ °K.

The electron temperature required to provide this value of brightness temperature is given by

$$T_b = T_e(1-e^{-\tau}),$$

where τ is the average, optical depth. The value of τ at 40 Mc/s. cannot be greater than unity, or the spectrum curve would fall below the lower limit of the experimental value shown. Substituting this value, $T_e = 3 \cdot 8 \times 10^7$ °K., which is a lower limit for the electron temperature.

Again, if Minkowski's (1942) estimate of the total thickness of the shell of the nebula, about 0·1 parsec, is accepted, an estimate of the electron density may also be found from the formula (e.g. Piddington, in press)

$$\tau = 0 \cdot 14 N^2 t / f^2 T_e^{3/2},$$

where N is the electron density, t the average thickness ($1 \cdot 2 \times 10^{18}$ cm.), f the wave frequency (40 Mc/s.). When τ has unit value, the value of N is 5×10^4 cm.$^{-3}$.

These lower limits of electron temperature and density (4×10^7 °K. and 5×10^4 cm.$^{-3}$ respectively) may be compared with *upper* limits of about 5×10^5 °K. and 2×10^3 cm.$^{-3}$ estimated by Minkowski from optical data. The discrepancy is large and it is concluded that if the origin is thermal, then existing ideas about the nebula will have to be radically revised.*

Mechanisms of radiation other than thermal have been suggested, depending on the plasma frequency (e.g. Martyn 1947) and on a differential velocity of two streams of ionized gas (e.g. Haeff 1949). The former would require electron densities as high as 10^{10} cm.$^{-3}$ to account for the higher frequency observations and may therefore be excluded. The latter seems a possible mechanism and would require differential velocities of a few kilometres per second between interpenetrating clouds of gas known to be moving at about 1300 km. sec.$^{-1}$ (Mayall 1937). The remarkably uniform flux density of radiation over the frequency range 40 to 1200 Mc/s., is consistent with a thermal origin in an optically thin distribution of ionized gas. It would appear, therefore, to provide some evidence against non-thermal mechanisms whose spectra, discussed by Bailey (1948), Haeff (1949), and others, show distinctly variable flux densities. In the case of the Crab Nebula, however, the uniform value may result from the combined effects of a large number of radiating regions having different, non-uniform, spectra.

V. Radiation from Other Sources

Radiation was observed from a number of other sources, the most interesting being the constellation of Cygnus, which will be the subject of a second paper. Other objects studied were some of the known radio discrete sources, outstanding nebulae, and the Moon.

---

* Since the above analysis was made Dr. R. L. Minkowski, in a personal communication which is acknowledged with thanks, has pointed out another difficulty in reconciling the thermally radiating model with other factors. The mass of the model, even if it were composed entirely of electrons and protons, would be about 600 solar masses. As other elements are likely to be present in considerable quantities the mass is probably much greater and unacceptably high in view of our knowledge of stellar masses.

(*a*) *Known Radio Sources*

Of three known radio sources investigated at 1210 Mc/s., only one, Centaurus-A (e.g. Stanley and Slee 1950) at R.A. 13 hr. 22 min., Dec. 42° 37′ S. or $l=274°$, $b=+20°$ (e.g. Stanley and Slee 1950) gave observable radiation. The observed flux density was $4 \cdot 1 \times 10^{-24}$ W. m.$^{-2}$ (c/s.)$^{-1}$, with an uncertainty somewhat greater than 20 per cent., perhaps about 30 per cent. The others, Virgo-A and Hercules-A, could not be detected and it may be assumed that their flux density of radiation was less than 1×10^{-24} W. m.$^{-2}$ (c/s.)$^{-1}$.

No attempt was made at 3000 Mc/s. to observe Centaurus-A and the other two could not be seen, so that their flux densities must be less than about 2×10^{-23} W. m.$^{-2}$ (c/s.)$^{-1}$.

Using data published by Stanley and Slee (1950) in the frequency range 60 to 160 Mc/s., and Shain (personal communication) at 18·3 Mc/s., the spectrum of Centaurus-A is shown in Figure 6. It is evident that the slope of the spectrum curve decreases with increasing frequency. The steep slope at the lowest frequency suggests that the flux density might reach large values at low frequencies. Observations below 10 Mc/s. would be of great interest.

It is of interest to note that whereas the radiation from Virgo-A has fallen off in a ratio greater than 12·5 as the frequency is raised from 100 Mc/s. (Stanley and Slee 1950) to 1210 Mc/s., that from Centaurus-A has only fallen off in a ratio of about 4·5. The corresponding ratios for the discrete sources in Sagittarius-Scorpius (see above), Taurus, and Cygnus (Piddington and Minnett, in press) are about 1·2, 0·9, and 6·5.

(*b*) *Outstanding Nebulae*

Attempts were made to detect radiation from the Great Nebula in Andromeda (M 31) and from the largest planetary nebula (NGC 7293). No radiation was observed, either at 1210 or 3000 Mc/s., so that upper limits of flux density may be assigned as follows (in units of W. m.$^{-2}$ (c/s.)$^{-1}$) :

	1210 *Mc/s.*	3000 *Mc/s.*
M 31	1×10^{-24}	2×10^{-23}
NGC 7293 ..	2×10^{-24}	2×10^{-23}

The larger value for NGC 7293 at 1210 Mc/s. is due to relatively bad conditions during the experiment.

Since the experiments were completed Brown and Hazard (1950) have reported radiation from M 31 at 158 Mc/s. The level observed was 4×10^{-25} W. m.$^{-2}$ (c/s.)$^{-1}$ so that, if the spectrum were similar to that of our own Galaxy, the anticipated level at 1210 Mc/s. would be (e.g. Piddington, in press) $2 \cdot 4 \times 10^{-25}$ W. m.$^{-2}$ (c/s.)$^{-1}$, which is well below the minimum observable level.

(*c*) *The Moon*

Although the measurement of lunar radiation does not fall strictly within the province of this paper, it does provide a rough calibration check of the 1200 and 3000 Mc/s. equipment, as well as being of some interest in its own right.

Radiation at 1210 Mc/s. was measured on September 22, 1950, when the lunar phase angle was +125°. The power flux density observed was $4\cdot 9\times 10^{-24}$ W. m.$^{-2}$ (c/s.)$^{-1}$, corresponding to an apparent disk temperature of 212 °K. the uncertainty being about 30 per cent. In observing the Moon it was necessary, after lining up the aerial beam on the Sun, to allow for the parallax resulting from the relative proximity of the Moon.

Observations at 3000 Mc/s. were made by the beam-swinging method on August 4, 1949, when the lunar phase angle was +120°. The power flux density observed was $3\cdot 9\times 10^{\ 23}$ W. m.$^{-2}$ (c/s.)$^{-1}$, corresponding to an apparent temperature of 215 °K., the uncertainty being perhaps 30 per cent.

These results may be compared with those of Piddington and Minnett (1949) at a frequency of 24,000 Mc/s. They found an apparent temperature (before correcting for emissivity) of 180 °K., when the phase angle was 120°, and an average value during the lunar cycle of 215 °K. The agreement with the former is well within the experimental error and it is doubtful if any significance should be attached to the rather higher values obtained at the lower frequencies.

VI. Acknowledgments

The authors are grateful to Mr. J. V. Hindman, Mr. C. F. Fryar, and Mr. G. H. Trensky for assistance with the equipment and during observations, to Mr. W. N. Christiansen for the loan of the 18-ft. by 16-ft. aerial, to Dr. J. L. Pawsey and Mr. C. A. Shain for help with the manuscript, and to Mr. B. Y. Mills and Mr. C. A. Shain for providing data prior to its publication.

VII. References

Allen, C. W., and Gum, C. S. (1950).—*Aust. J. Sci. Res.* A **3** : 224.

Baade, W. (1942).—*Astrophys. J.* **96** : 188.

Bailey, V. A. (1948).—*Aust. J. Sci. Res.* A **1** : 351.

Bolton, J. G., and Stanley, G. J. (1949).—*Aust. J. Sci. Res.* A **2** : 139.

Bolton, J. G., and Westfold, K. C. (1950*a*).—*Aust. J. Sci. Res.* A **3** : 19.

Bolton, J. G., and Westfold, K. C. (1950*b*).—*Aust. J. Sci. Res.* A **3** : 251.

Brown, R. H., and Hazard, C. (1950).—*Nature* **166** : 901.

Dicke, R. H. (1946).—*Rev. Sci. Instrum.* **17** : 268.

Friis, H. T., and Feldman, C. B. (1937).—*Proc. Inst. Radio Engrs. N.Y.* **25** : 841.

Haeff, A. V. (1949).—*Phys. Rev.* **75** : 1546.

Hey, J. S., Parsons, S. J., and Phillips, J. W. (1948).—*Proc. Roy. Soc.* A **192** : 425.

Lehany, F. J., and Yabsley, D. E. (1949).—*Aust. J. Sci. Res.* A **2** : 48.

Martyn, D. F. (1947).—*Nature* **159** : 26.

Mayall, N. U. (1937).—*Publ. Astr. Soc. Pacif.* **49** : 104.

Minkowski, R. (1942).—*Astrophys. J.* **96** : 199.

Northcott, Ruth, and Williamson, R. E. (1948).—*J. Roy. Astr. Soc. Canad.* **42** : 169.

Piddington, J. H.—*Mon. Not. R. Astr. Soc.* The origin of galactic radio-frequency radiation (in press).

Piddington, J. H., and Hindman, J. V. (1949).—*Aust. J. Sci. Res.* A **2** : 524.

Piddington, J. H., and Minnett, H. C. (1949).—*Aust. J. Sci. Res.* A **2** : 63.

Piddington, J. H., and Minnett, H. C.—*Aust. J. Sci. Res.* A **5**. Radio-frequency radiation from the constellation of Cygnus (in press).

Reber, G. (1944).—*Astrophys. J.* **100** : **279**.

REBER, G. (1948).—*Proc. Inst. Radio Engrs. N.Y.* **36** : 1215.
RYLE, M., and SMITH, F. G. (1948).—*Nature* **162** : 462.
RYLE, M., and VONBERG, D. D. (1948).—*Proc. Roy. Soc.* A **193** : 98.
SEEGER, C. L., and WILLIAMSON, R. E. (1951).—*Astrophys. J.* **113** : 21.
STANLEY, G. J., and SLEE, O. B. (1950).—*Aust. J. Sci. Res.* A **3** : 234.
TOWNES, C. H. (1947).—*Astrophys. J.* **105** : 235.

GALACTIC RADIATION AT RADIO FREQUENCIES

IV. THE DISTRIBUTION OF RADIO STARS IN THE GALAXY

By J. G. Bolton* and K. C. Westfold*

[*Manuscript received April* 26, 1951]

Summary

The distribution of sources of galactic noise at 100 Mc/s. in the galactic plane is deduced from the data of a previous survey, assuming axial symmetry. Effects of absorption and emission within the interstellar gas are considered and found to be negligible. Away from the galactic plane, the noise contours do not conform to a single spheroidal source distribution. Consideration is given to an isotropic contribution to the general background radiation—possibly extragalactic in origin.

On the hypothesis that the background radiation is the aggregate emission from a distribution of radio stars typified by the discrete sources previously studied, estimates of the local number density, the flux from a typical radio star, and the distances of some of the observed discrete sources are made. These estimates are briefly discussed.

I. Introduction

In Part III of this series (Bolton and Westfold 1950*b*) it was recalled that neither the mechanism of free-free transitions in the interstellar gas, nor the aggregate emission from stars behaving like the Sun at times of high spot activity, could account for the intensity of the general background of galactic radiation. In the present paper we adopt the hypothesis that the background galactic radiation is the aggregate emission from the distribution of a new type of star, which is typified by the discrete sources studied extensively by Bolton, Ryle, Lovell, and their co-workers. There is as yet no proof that the known discrete sources are galactic objects. Two of them have, in fact, been provisionally identified with extragalactic nebulae (M 87 and NGC 5128—Stanley and Slee 1950 ; Part II of this series). Although the positions of a number of them are known with a fair degree of accuracy, not one has been identified with any outstanding visible star—a feature which makes it difficult to speculate profitably on the optical type of the discrete sources and their means of energy generation. At this stage we shall refer to the new type stars as *radio stars*, endowing them with the property of high radio and low optical emission.

In Part III, the results of our 100 Mc/s. survey (Bolton and Westfold 1950*a* ; Part I of this series) were interpreted as indicating that the Sun lies in or near an arm of a spiral galaxy. This interpretation rested on the simple qualitative hypothesis that the intensity of the galactic radiation in any direction is a measure of the extent of the Galaxy in that direction. It contains the implication that the radio stars are distributed over more or less the same regions of space as the visible stars.

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

In the present paper* the same 100 Mc/s. data are analysed in more detail; by assuming axial symmetry about the centre of the Galaxy, the distribution of sources in the galactic plane is deduced as a function of the distance from the centre almost out to the Sun's distance. In Appendix I it is shown that the contours of the radiation from outside the galactic plane do not conform to a single spheroidal distribution, a possibility suggested by the form of the contours.

Before carrying out this analysis, the findings of recent studies on the effects of absorption and emission by interstellar gas were examined. With limited exceptions these are negligible at 100 Mc/s.

Finally, some inferences which can be drawn from the identification of the radio stars producing the general background with the known discrete sources are discussed. Rough estimates are derived for the flux emitted by an average radio star and the local density of the radio stars.

II. INTERSTELLAR GAS

The physical conditions in interstellar space have been the subject of detailed exploratory investigations by Strömgren (1948), Spitzer (1948, 1949), and Spitzer and Savedoff (1950). It appears likely that 90 per cent. of the space is occupied by ionized H II regions where the kinetic temperature is about 10,000 °K. and the electron and hydrogen densities are about 0·1 cm.$^{-3}$. The remainder of the space is occupied by interstellar clouds. Most clouds are H I regions where the temperature probably lies between 1000 and 100 °K., the preponderant atomic population being hydrogen of density between 15 and 5 cm.$^{-3}$, with corresponding electron density between 0·01 and 0·004 cm.$^{-3}$. Clouds about O and early B stars are H II regions with electron and hydrogen densities of up to 100 cm.$^{-3}$ and temperatures up to 13,000 °K. These are, however, comparatively rare and will be considered later.

TABLE 1

Region	Kinetic Temperature (°K.)	Electron Density (cm.$^{-3}$)	Hydrogen Density (cm.$^{-3}$)	Absorption Coefficient for 100 Mc/s. Radiation (parsec^{-1})	Optical Thickness of 20,000 parsecs
H II	10,000	0·1	0·1	$3 \cdot 6 \times 10^{-7}$	$7 \cdot 2 \times 10^{-3}$
H I(*a*)	1000	0·01	15	$1 \cdot 05 \times 10^{-9}$	$(2 \cdot 1 \times 10^{-4})$
H I(*b*)	100	0·004	5	$4 \cdot 4 \times 10^{-10}$	$(8 \cdot 8 \times 10^{-6})$

The values of the absorption coefficients of such regions for 100 Mc/s. radiation are set out in Table 1. For the H II regions the formula given by Burkhardt, Elwert, and Unsöld (1948) with the logarithmic factor equal to 12, gives

---

* A summary of the material in this paper was given at the 1950 meeting of U.R.S.I. at Zurich.

$$\varkappa_f = 3 \cdot 6 \times 10^{17} (nNT^{-3/2}/f^2) \text{ parsec}^{-1},$$

where n and N are the electron and hydrogen densities in cm.$^{-3}$, f the frequency in sec.$^{-1}$, and T the kinetic temperature in °K. For the H I regions the ordinary free-path Lorentz formula in terms of the collision frequency of electrons with atoms gives, taking the value 3×10^{-8} cm. for the mean collision diameter as suggested by Chapman and Bartels (1940),

$$\varkappa_f = 2 \cdot 2 \times 10^{17} (nNT^{-\frac{1}{2}}/f^2) \text{ parsec}^{-1}.$$

It can be seen that the H II regions have the greatest effect on the radiation. We have added corresponding values of the optical thickness of 20,000 parsecs, the greatest extent of the Galaxy from the Sun through the centre. Since the maximum optical depth that could be traversed by a 100 Mc/s. ray is $7 \cdot 2 \times 10^{-3}$ parsecs, we are entitled to neglect absorption effects entirely in the following considerations.

The figures also indicate the inadequacy of the hypothesis that the background galactic radiation originates in free-free transitions in interstellar gas. In such an optically thin medium, the contribution to the *brightness temperature** of a ray is the product of the optical depth and the kinetic temperature, which in this case amounts to no more than 70 °K. Other objections, pointed out by Unsöld (1949*a*, 1949*b*) are the relatively low concentrations of the contours about the galactic plane and the spectrum of the radiation. In the region of the galactic poles the intensity varies as $\lambda^{0 \cdot 35}$ (Part II of this series) in the metre wavelength range, and indications are that the index is higher for regions nearer the galactic plane. If the radiation were due to free-free emission, the index would be zero or negative.

It is pointed out that in H II clouds associated with O and B stars, the absorption coefficient and the optical depth may be increased by a factor as large as 10^6, and a cloud of more than 10 parsecs in depth would contribute its full kinetic temperature. Development of more sensitive receiving equipment should ultimately permit detection of radiation from some of the nearer clouds. Considering observations over a finite solid angle, the effect of such clouds in emission or absorption will be governed by the dilution factor for the clouds within that solid angle. This will only be appreciable looking along a spiral arm of the Galaxy—in the Cygnus or Carina directions perhaps. In the general direction of the galactic centre with which we are mainly concerned in the later sections of the present paper, the effect of the H II clouds may be safely ignored. In certain directions in Cygnus, the dilution factor (i.e. over several square degrees) may be as high as $0 \cdot 2$—some idea of it may be gained from the corresponding dilution factor for the strong Hα emission, as suggested by Woolley (1947).

---

* In accordance with the recommendation of the Commission V Subcommittee on Radio-Astronomy Nomenclature at the 1950 U.R.S.I. meeting, we have adopted the term *brightness temperature* in place of *equivalent temperature* used in earlier papers.

III. The Source Distribution in the Galactic Plane

The problem of determining the distribution of the sources of galactic noise in any plane through the Sun from the noise profile, depends essentially on the prior knowledge of the curves of constant source density in that plane. The only reasonable assumption of this character that can be made is that the distribution is symmetrical about the axis of rotation of the Galaxy almost out to the Sun's distance, where the spiral structure would presumably destroy the symmetry. Then the curves of constant source density in planes perpendicular to the axis are circles. In particular, the noise profile in the galactic plane can be analysed on the basis of a radial source distribution.

The profiles through the galactic centre in and perpendicular to the galactic plane are shown in Figure 1, in terms of the brightness temperature. The broken line at 1000 °K. separates, somewhat arbitrarily, the contribution assumed to be due to " extranuclear " sources. The boundary imposed would seem to be quite reasonable considering the fairly sharp cut-offs in the noise profiles near $l'=\pm 80°$ and $b'=\pm 20°$, though alternative forms of the boundary line could have been adopted. However, since the assessment of the form of the extranuclear contribution would have been a pure guess, the simplest form of separation was chosen. The following analysis depends principally on the gradients of the noise profiles and not on the absolute values, so that only small errors are introduced by this approximation.

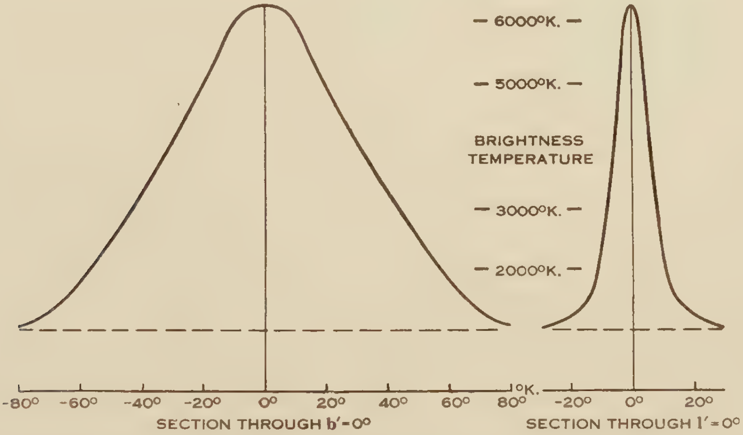


Fig. 1.—Profiles of 100 Mc/s. noise distribution in and perpendicular to the galactic plane.

The coordinates with respect to the noise centre are

$$\left.\begin{aligned} l' &= l - l_0, \\ \text{and} \quad b' &= b - b_0, \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

where l and b are the galactic coordinates of Part I, and ($l_0=326°$, $b_0=-1°$) is the position of the noise centre. The ordinates are the mean values of brightness temperatures in directions symmetrical about this centre.

The above values of l_0 and b_0 represent a new direct determination of the position of the galactic centre. They are subject to a small precession correction of 1°·5 in right ascension, due to using epoch 1900 charts for converting from the equatorial coordinates of the 1949 observations to galactic coordinates. Making the necessary corrections, the coordinates of the centre are

$$\begin{aligned} l_0 &= 326^\circ{\cdot}4 \pm 2^\circ{\cdot}0, \\ b_0 &= -0^\circ{\cdot}4 \pm 0^\circ{\cdot}5. \end{aligned} \qquad 1900$$

This places the Sun virtually in the galactic plane.* The result is in good agreement with van Tulder's (1942) discussion of the position of the galactic pole from analysis of the motions of visible stars. The corresponding position for the centre is

$$\begin{aligned} l_0 &= 325^\circ{\cdot}9 \pm 0^\circ{\cdot}3, \\ b_0 &= 0^\circ{\cdot}6 \pm 0^\circ{\cdot}3. \end{aligned} \qquad 1900$$

The procedure for analysing the " nuclear " contribution to the noise profile in the galactic plane is essentially an extension of the classical method of determining the distribution of stars in globular clusters (cf. Smart 1938) to the case where the rays are not parallel. The brightness temperature T_e of a ray from any direction (l', b') is the sum of contributions from the individual radio stars along its path. If at a distance R from the Sun the density of radio stars of average disk area A and monochromatic disk temperature T_s is n, the contribution from the element of length dR is $nAT_s dR$. This result may be regarded as the consequence of a differential dilution factor $nAdR$, or of a volume emissivity of $(k/c^2)f^2nAT_s$, where k is Boltzmann's constant, c the velocity of light, and f the frequency of the radiation.

Then, generally,

$$T_e(l', b') = \int_0^{R(l',b')} nAT_s dR, \quad \dots\dots\dots\dots\dots \quad (2)$$

where $R(l',b')$ is the total linear extent of the regions occupied by sources in the direction (l',b'). If this relation is restricted to contributions from the " nucleus " in the galactic plane, the source density function nAT_s is a function of the distance from the centre only. Thus

$$nAT_s = F(\xi) \quad \dots\dots\dots\dots\dots\dots\dots \quad (3)$$

say, where ξ is the square of the distance from the centre of the point whose galactic coordinates are $(R,l',0)$ (Fig. 2). Taking the distance from Sun to centre as unity,

$$\xi = 1 - 2R \cos l' + R^2. \quad \dots\dots\dots\dots\dots\dots \quad (4)$$

Along any path $(l',0)$ from the Sun, ξ decreases from its maximum value Ξ, which limits the " nucleus ", to its minimum value $\eta(l',0)$ (where $\eta = \sin^2 l'$) after which it increases again to Ξ. Then from (2) and (3)

$$T_e(l',0) = \int_{\eta(l',0)}^{\Xi} F(\xi)\left(\frac{\partial R_1}{\partial \xi} - \frac{\partial R_2}{\partial \xi}\right) d\xi, \quad \dots\dots\dots\dots\dots \quad (5)$$

---

* Recently Seeger and Williamson (1951) have made a determination of the position of the galactic pole from noise profiles through the galactic equator at various longitudes. Their position of the pole is appreciably different from van Tulder's, but their work is subject to the criticism that the observations used did not cover between longitudes 100 and 330°. Away from the galactic centre results may be seriously influenced by local structure.

where $R_1(\xi,l',0)$, $R_2(\xi,l',0)$ are the distances of farther and nearer intersections of the paths with the circle ξ=constant (Fig. 2). From (4)

$$\left.\begin{aligned} R_1&=\sqrt{(1-\eta)}+\sqrt{(\xi-\eta)},\\ R_2&=\sqrt{(1-\eta)}-\sqrt{(\xi-\eta)}.\end{aligned}\right\} \quad \ldots\ldots\ldots\ldots \quad (6)$$

Hence

$$\left.\begin{aligned} T_e(l',0)&=\int_\eta^\Xi F(\xi)\frac{d\xi}{\sqrt{(\xi-\eta)}}\\ &=G(\eta),\ \text{say.}\end{aligned}\right\} \quad \ldots\ldots\ldots\ldots \quad (7)$$

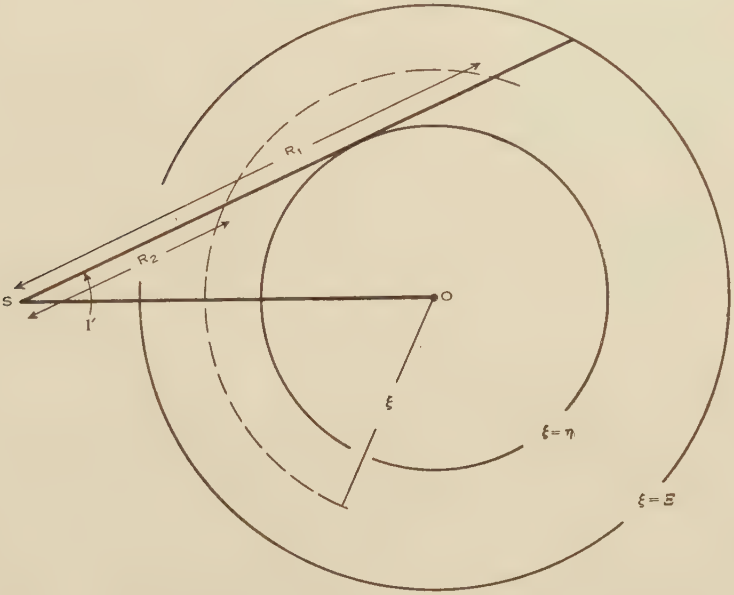


Fig. 2.—Ray path through the nucleus to the Sun.

The representation of the function $G(\eta)$ is obtained by transforming the left-hand profile of Figure 1 into a plot against $\sin^2 l'$.

The relation (7) is the well-known Abelian integral equation that occurs in the analogous analysis for the globular clusters. Its solution is

$$\begin{aligned} F(\xi)&=-\frac{1}{\pi}\int_\xi^\Xi G'(\eta)\frac{d\eta}{\sqrt{(\eta-\xi)}}\\ &=-\frac{2}{\pi}(\Xi-\xi)G'(\Xi)+\frac{2}{\pi}\int_\xi^\Xi\sqrt{(\eta-\xi)}G''(\eta)d\eta. \quad \ldots\ldots\ldots \quad (8)\end{aligned}$$

The latter form is more suitable for numerical integration. The functions $G(\eta)$, $G'(\eta)$, and $G''(\eta)$ are plotted in Figure 3. It can be seen that the gradient of $G(\eta)$ is constant from $\eta=0\cdot34$ to $\eta=0\cdot8$. To simplify the numerical work this linear part was extrapolated to meet the η-axis, thus making the limiting value for the nucleus $\eta=\Xi=0\cdot9$. This modification is of no great significance as the nucleus has already been delimited in a somewhat arbitrary manner. The results of the analysis are shown in Figure 4, which depicts $F=nAT_s$ as a function of the radial distance $\sqrt{\xi}$. The sharp cut-off at the limit $\sqrt{\Xi}=0\cdot95$ is due to the extrapolation of the $G(\eta)$ curve. It should be emphasized that the curve does not necessarily represent the density distribution of radio stars, but only of the source function nAT_s. As with the visible stars, there may be several types of radio star, each with its characteristic spectrum and distribution in space. The resolution of such possible types and their distribution requires further investigations on the spectra of the discrete sources and on the distribution of the background galactic radiation over a wide range of frequencies.

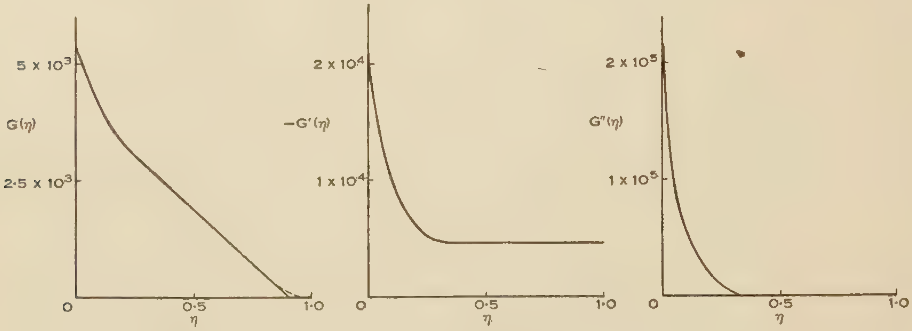


Fig. 3.—The functions $G(\eta)$, $G'(\eta)$, and $G''(\eta)$.

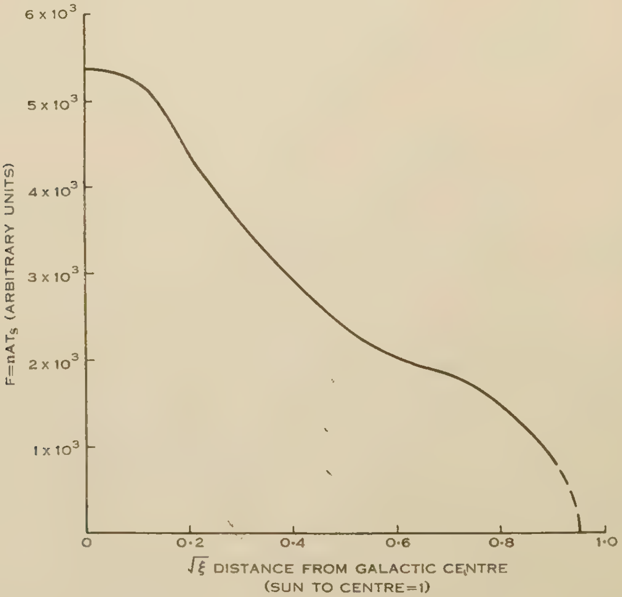


Fig. 4.—" Nuclear " source distribution in the galactic plane.

In Appendix I the analysis of this section is extended beyond the galactic plane to a single spheroidal source distribution. It is found that the right-hand profile of Figure 1 does not conform to such a model. However, the possibility still remains that the radio stars are distributed in a number of spheroidal systems with different degrees of flattening determined by the velocity of the rotation, as suggested by Lindblad (1948) for the visible stars. Without definite knowledge it seems practically an impossible task to find by trial the forms of the curves F = constant in this and other planes. However, the analysis of this section can be applied to planes of constant latitude b' up to about 5° wherein nuclear sections will still be approximately circular.

The chief feature of the source distribution of Figure 4 is the relatively high concentration about the centre, up to one-half the distance to the Sun. In the outer regions there is no trace of a "ripple" which could be associated with spiral structure within the "nucleus". This should not be expected, firstly, because of the low resolving power of the aerial system used in the survey, and secondly, because such structure is unlikely to be radially symmetrical.

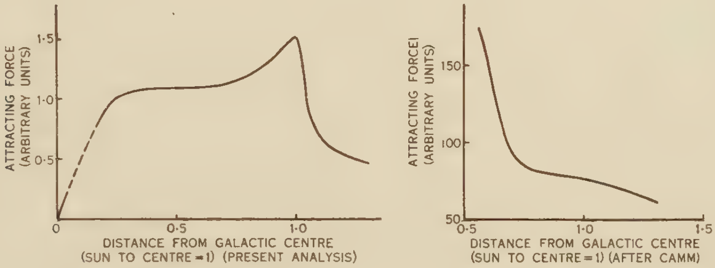


Fig. 5.—The variation of gravitational attraction in the galactic plane.

Although it is unlikely that the radio stars comprise the bulk of the galactic mass, it has been suggested (e.g. Unsöld 1949*b*) that the noise contours about the galactic centre represent a distribution similar to that of the stars in Baade's Population II. As Population II stars probably form the bulk of the galactic mass, it is of interest to determine the gravitational attracting force in the galactic plane for a density distribution having the shape of the source distribution of Figure 4. As a rough approximation we assume that the mass is concentrated into a thin disk, and follow the procedure adopted by Wyse and Mayall (1942). The result is shown in the left-hand diagram of Figure 5—it appears to be a hybrid of the models in Figures 2 and 4 of Wyse and Mayall.* The right-hand diagram of Figure 5 is the gravitational attraction deduced by Camm (1938) from a study of the radial velocities of planetary nebulae. The two curves have the same form over a restricted range, but the detailed agreement is very poor. However, Camm admits that his results are uncertain for distances less than 0·8 of the distance from the centre of the Galaxy to the Sun, and our own analysis cannot be taken seriously at distances greater than 0·8 owing to the approximations previously mentioned.

---

* This can be readily understood if the two halves of the right-hand side of equation (8) and the functions involved are examined.

IV. The Discrete Sources and the General Background

Before the data provided by studies of the discrete sources are considered in relation to the general background of radiation, it seems worthwhile discussing the possible existence of an isotropic contribution. From equation (2) of the previous section it follows that the average value of the source density function $F=nAT_s$ in any direction is T_e/R. For the direction of the galactic centre this is 0·38 °K. parsec^{-1}, taking a value of 8000 parsecs for the distance from the Sun to the galactic centre. In the direction of the regions of minimum noise brightness near the galactic poles, the brightness temperature is about 500 °K. The extent of the Galaxy in these directions can hardly be greater than 1000 parsecs, giving an average value of $F=0\cdot5$ °K. parsec^{-1}, i.e. higher than that for the direction of the centre. It is in fact almost as high as the value of 0·67 °K. at the galactic centre. (The units for F in Figure 4 are arbitrary, being independent of the actual distance from the Sun to the galactic centre. The figure of 0·67 is obtained by dividing the value of F in arbitrary units by 8000—the distance from the Sun to the centre in parsecs.) Additional evidence for an isotropic background is to be found in the low contrast between the brightness temperatures of 500 °K. near the poles and the temperature in the galactic plane but away from the centre. At longitude 210° the temperature is only 650 °K., with values ranging up to 1000 °K. in the subsidiary maximum at longitude 150°. Although uncertainties in the extent of the Galaxy in different directions would make quantitative estimates unreliable, the evidence presented suggests that the observed noise distribution represents the superposition of two components. The first of these is to be associated with the distribution of radio stars discussed in the previous section, and the second an essentially isotropic background. This isotropic background may be extragalactic in origin, or may even originate in the solar system in a manner similar to that proposed by Alfvén and Herlofsen (1950) in their radio star model, discussed later.

We now examine the consequences of identifying the known discrete sources with the hypothetical radio stars of the last section. We rewrite equation (3) in the form

$$F=nAT_s, \qquad \text{(9)}$$

and add to it information gained from observational studies of the discrete sources. If p_{min} per unit bandwidth is the minimum flux that can be detected by the receiving system,

$$p_{min}=\frac{k}{c^2}f^2\frac{AT_s}{r_{max}}, \qquad \text{(10)}$$

where r_{max} is the maximum distance at which an average source could be detected. The total number of sources capable of detection, per unit solid angle in any direction, is

$$N=\tfrac{1}{3}r^3_{max}n. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \mathbf{(11)}$$

Equations (**9**), (**10**), and (**11**) give

$$\left.\begin{aligned} r_{max}&=3N\,\frac{1}{F}\cdot\frac{p_{min}}{kf^2/c^2},\\ n&=\frac{1}{(3N)^2}F^3\left(\frac{kf^2/c^2}{p_{min}}\right)^3,\\ AT_s&=(3N)^2\frac{1}{F^2}\left(\frac{p_{min}}{kf^2/c^2}\right)^3,\end{aligned}\right\} \quad \ldots\ldots \quad \mathbf{(12)}^*$$

in terms of known quantities. For the 100 Mc/s. equipment at Dover Heights (Parts I and II of this series), $p_{min}=10^{-24}$ W.m. ^{2}(c/s.)$^{-1}$.† The total number of sources detectable over the whole celestial sphere with this equipment is not exactly known. Some thirty sources have been discovered in the southern hemisphere, but the actual number may be greater because of the possibility of the interference patterns of weak sources being masked by the patterns of more intense adjacent sources. Ryle (1950) reports some thirty sources down to approximately the same limits of sensitivity. As a conservative estimate and as an average over the whole celestial sphere we take $4\pi N=50$. A larger value will increase r_{max} and AT_s and decrease n. The choice of a suitable value for F is a major difficulty. A reasonable assumption is that F is less than $0\cdot2$ °K. parsec^{-1}—the value at $\sqrt{\xi}=0\cdot5$ where the distribution in Figure 4 shows signs of flattening out. Adopting this value (**12**) gives

$$\begin{aligned} r_{max}&\geq 40 \text{ parsecs},\\ n&\leq 2\times10^{-4} \text{ parsec}^{-3},\\ AT_s&\geq 10^3 \text{ °K. parsec}^2.\end{aligned}$$

The low value for r_{max} justifies to some extent the choice of a value for F which is lower than the average for the direction of the galactic centre. In the circumstances a still lower value might well be adopted, but it is pointed out that such a procedure could not be continued indefinitely, as r_{max} would increase until a point was reached where the low value of F would no longer hold. A decrease in F has the same effect as an increase in N. In both cases n and AT_s are more sensitive to changes than r_{max}.

If the present lower limit for r_{max} is of the right order of magnitude, the observed discrete sources would be local objects—that is on a galactic scale.

---

[*] This analysis could, of course, be extended into the usual form, in terms of the number N_p of sources per unit solid angle contributing a flux greater per unit bandwidth than p, throughout the range of p. However, the total number of sources so far discovered is too small for such an analysis to be significant.

[†] This sensitivity refers to both polarizations, whereas equation (**10**) refers to one polarization only. However, the 30 sources mentioned have intensities greater than 10^{-24} W.m.$^{-2}$ (c/s.)$^{-1}$ for one polarization.

The value for n* indicates that they would be fairly rare as the local star density is about 0·1 per cubic parsec. The value for AT_s may be compared with that for the Sun at similar frequencies—2×10^{-9} parsec^2 °K., assuming a disk area of 2×10^{-15} parsec^2 and an effective temperature of 10^6 °K. under " quiet Sun " conditions. During intense solar outbursts the effective temperature has been observed to attain values of the order of 10^{13} °K., but such occurrences are rare and the outbursts only last for a period ranging from several minutes to an hour· The radiation from the discrete sources is incomparably greater than the average solar output.

The high value† of AT_s for the radio stars may be realized by adopting large effective disk areas. Alfvén and Herlofsen (1950) have recently pointed out that, according to the theory of Teller and Richtmyer, the Sun should be surrounded by a magnetic field of the order of 10^{-6} to 10^{-5} gauss, which is sufficient to confine cosmic radiation to within a region of dimension about 0·1 light-year. If this is so, it is likely that other stars have their own cosmic radiation likewise confined. These authors calculate that high-energy cosmic ray particles moving in such fields would radiate mainly in the radio spectrum. To give the required flux a radio star should not emit very much light, and should be situated within an interstellar gas cloud. For such a diameter the disk temperature would be only 10^6 °K. It is worthwhile to note in this connection that if the value of r_{max} is adopted and the source in Cygnus is an average source, its distance would be only of the order of 3·5 parsecs. The experimental upper limit to the angular diameter for this source is 1·5 minutes of arc, which would give an actual diameter of 0·005 light-years. Such considerations do not apply in the case of the other less intense sources, for which the experimental upper limits are much higher and the probable distances much greater. It does, however, emphasize the need for obtaining improved data on angular widths of the sources.

---

* Our result for n is much smaller than the figure of 3 parsec^{-3} quoted by Ryle (1950) from an analysis by Smith. Their revised figure of 0·45 parsec^{-3} (personal communication) is smaller, but still greater than the local density of visible stars. Smith's analysis is of the type described in the footnote to equation (**12**). The plot of $p_N^{-3/2}$ against N, where p_N is the flux from the Nth most intense source, should be a straight line through the origin, whose gradient determines $(T_e/R)^{-3/2}n^{\frac{1}{2}}$. The points representing the individual sources lie very close to a straight line, but this does not pass through the origin. However, apart from the neglect of possible extra-galactic contributions, the chief source of the discrepancy appears to be in the value adopted for R. Assuming that the distribution of the discrete sources is similar to that of the visible stars in the observed strip of the sky, Smith deduces R as the radius of an equivalent spherical galaxy centred at the Sun. He thus finds $R=1000$ parsecs, which seems to be too small a figure for the average galactic extent over this region.

† The above values of n and AT_s can be considerably modified by the assumption of a reasonable dispersion in the values of AT_s. In a paper to be published in B.A.N. 426, Westerhout and Oort have shown that this can raise n from $3\cdot2\times10^{-4}$ to 0·27 parsec^{-3} and reduce AT_s from 260 to a spread about 5 °K. parsec^2.

V. References

Alfvén, H., and Herlofsen, N. (1950).—*Phys. Rev.* **78** : 616.
Bolton, J. G., and Westfold, K. C. (1950*a*).—*Aust. J. Sci. Res.* A **3** : 19-33.
Bolton, J. G., and Westfold, K. C. (1950*b*).—*Aust. J. Sci. Res.* A **3** : 251-64.
Burkhardt, G., Elwert, G., and Unsöld, A. (1948).—*Z. Astrophys.* **25** : 310-4.
Camm, G. L. (1938).—*Mon. Not. R. Astr. Soc.* **99** : 71-90.
Chapman, S., and Bartels, J. (1940).—" Geomagnetism." Vol. 1. (Oxford Univ. Press.)
Lindblad, B. (1948).—*Mon. Not. R. Astr. Soc.* **108** : 214-35.
Ryle, M. (1950).—*Rep. Progr. Phys.* **13** : 184-246.
Seeger, C. L., and Williamson, R. E. (1951).—*Astrophys. J.* **113** : 21-49.
Smart, W. M. (1938).—" Stellar Dynamics." (Cambridge Univ. Press.)
Spitzer, L. (1948).—*Astrophys. J.* **107** : 6-33.
Spitzer, L. (1949).—*Astrophys. J.* **109** : 337-53.
Spitzer, L., and Savedoff, M. P. (1950).—*Astrophys. J.* **111** : 593-608.
Stanley, G. J., and Slee, O. B. (1950).—*Aust. J. Sci. Res.* A **3** : 234-50.
Strömgren, B. (1948).—*Astrophys. J.* **108** : 242-75.
van Tulder, J. J. M. (1942).—*Bull. Astr. Insts. Netherlds.* **6** : 315-25.
Unsöld, A. (1949*a*).—*Nature* **163** : 489-91.
Unsöld, A. (1949*b*).—*Z. Astrophys.* **26** : 176-99.
Woolley, R. v. d. R. (1947).—*Mon. Not. R. Astr. Soc.* **107** : 308-15.
Wyse, A. B., and Mayall, N. U. (1942).—*Astrophys. J.* **95** : 24-48.

Appendix I

The Analysis of a Spheroidal Source Distribution

If a right-handed rectangular coordinate system $Oxyz$ through the centre of the Galaxy is taken as Figure 6, so that Oz is in the direction of the north galactic pole and Ox in the direction from centre to Sun, the galactic coordinates R, l', b' are related to x, y, z, so that

$$\left.\begin{aligned} x&=1-R\cos l'\cos b',\\ y&=1-R\sin l'\cos b',\\ z&=R\sin b'.\end{aligned}\right\} \quad \ldots\ldots\ldots\ldots \quad \textbf{(A1)}$$

Then if the axially symmetrical source distribution is spheroidal,

$$nAT_s=F(\xi), \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \textbf{(A2)}$$

where

$$\xi=x^2+y^2+k^2z^2. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad \textbf{(A3)}$$

The ratio of the axes, k, is greater than unity if the spheroids are oblate.

Then, as in Section III, the brightness temperature of a ray from the direction (l',b') is

$$T_e(l',b')=\int_{\eta(l',b')}^{\Xi}F(\xi)\left(\frac{\partial R_1}{\partial\xi}-\frac{\partial R_2}{\partial\xi}\right)d\xi, \quad \ldots\ldots\ldots \quad \textbf{(A4)}$$

where now

$$\left.\begin{aligned} R_1(\xi,l',b')&=\frac{\sqrt{(1-\eta)}+\sqrt{(\xi-\eta)}}{\sqrt{(\cos^2 b'+k^2\sin^2 b')}},\\ R_2(\xi,l',b')&=\frac{\sqrt{(1-\eta)}-\sqrt{(\xi-\eta)}}{\sqrt{(\cos^2 b'+k^2\sin^2 b')}},\\ \eta(l',b')&=1-\frac{\cos^2 l'\cos^2 b'}{\cos^2 b'+k^2\sin^2 b'}.\end{aligned}\right\}. \quad \ldots\ldots \quad \textbf{(A5)}$$

In the galactic plane $b'=0$ and (**A5**) reduces to (**6**). Instead of (**7**) we get

$$\sqrt{(\cos^2 b'+k^2 \sin^2 b')}T_e(l',b')=\int_\eta^\Xi F(\xi)\frac{d\xi}{\sqrt{(\xi-\eta)}}, \quad \ldots\ldots \quad \textbf{(A6)}$$
$$=G(\eta),$$

where $G(\eta)$ is the same function as in Section III. Thus, if the source distribution is spheroidal, any plot of $\sqrt{(\cos^2 b'+k^2 \sin^2 b')}T_e(l',b')$ against $\eta(l',b')$ will be identical with the plot of $G(\eta)$ in Figure 3.

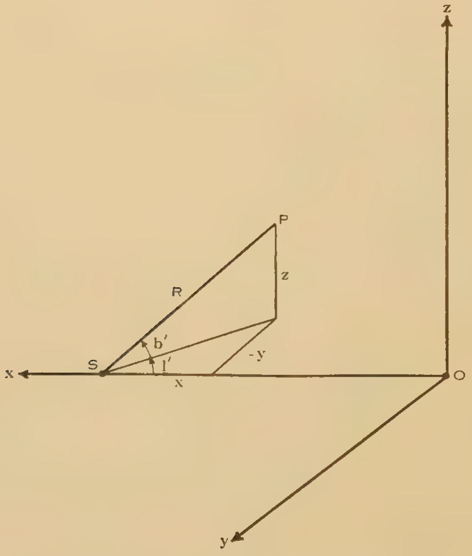


Fig. 6.—The galactic coordinates (R,l',b') and fundamental coordinates (x,y,z).

So far, k is undetermined. It may be found from the identification of both sides of (**A6**) for a single direction (l',b'). Then the comparison of both sides for other directions constitutes a direct test of the hypothesis of a spheroidal source distribution. The direction (0, 10°) was chosen to determine k, giving $k=5\cdot8$. Then both sides of (**A6**) were compared at other points on the right-hand profile of Figure 1. No agreement was found, the values of $\sqrt{(\cos^2 b'+k^2 \sin^2 b')}T_e(0,b')$ being progressively greater and less than $G(\eta)$ according as b' is less than and greater than 10°.

THE POSITION AND MOVEMENT ON THE SOLAR DISK OF SOURCES OF RADIATION AT A FREQUENCY OF 97 Mc/s.

I. EQUIPMENT

By A. G. Little* and Ruby Payne-Scott*

[*Manuscript received July 3, 1951*]

Summary

The spaced aerial interference technique, as used in radio astronomy, has been modified in order to observe solar disturbances which are either of short duration or are rapidly changing in position. The interference pattern is obtained in 1/25 of a second by sweeping the aerial lobe pattern across the source, 25 times per second. The position of the source and the polarization of the radiation are determined separately, a complete record taking about one second. Two different aerial spacings are used and calibration is carried out by a novel method.

An equipment operating on 97 Mc/s., which has been used for observations of solar disturbances, is described and the sources of error discussed.

I. Introduction

In recent years, numerous observations have been made of the radiation emitted by the Sun at metre wavelengths. Appleton and Hey (1946*a*), Pawsey, Payne-Scott, and McCready (1946), and McCready, Pawsey, and Payne-Scott (1947) have found that this radiation has two components : a steady " thermal component " and a " variable " component which, it is assumed, is non-thermal in origin. This latter is not present at all times, and constitutes a very great increase in intensity above the level of the thermal radiation. As shown by Payne-Scott (1949), this increase can take either of two forms : a slowly-varying long-term " noise storm " or a short-duration, high-intensity burst, or outburst. Both these increases may occur together, but it quite often happens that one occurs without the other.

McCready, Pawsey, and Payne-Scott (1947) have shown that the radiation during a noise storm originates within a restricted area, near sunspots, and measurements by Ryle and Vonberg (1946), Martyn (1946), and Appleton and Hey (1946*b*) show that it is at most times circularly polarized.

Very little information as to the origin of bursts has been available, although observations by Appleton and Hey (1946*a*), Lovell and Banwell (1946), Payne-Scott (1946), Allen (1947), and Covington (1948) have shown a correlation between outbursts and solar flares, whilst Payne-Scott (1949) has shown that some bursts are not circularly polarized. These measurements, however, have not covered all aspects of the variable components and further observations of their place of origin and polarization are necessary to obtain a complete picture of the metre-wavelength radiation.

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

For accurate position measurements, an aerial having the required resolving power becomes prohibitively large, and as a result interference techniques have been developed which allow small, wide-beam aerials to be used. McCready, Pawsey, and Payne-Scott (1947) and Bolton and Stanley (1948*a*, 1948*b*) have used this technique in the form of a Lloyd's mirror interferometer, in which interference is established between the direct and reflected rays from a source rising over the sea. Ryle and Vonberg (1946, 1948), have used a radio version of Michelson's stellar interferometer, employing two horizontally spaced aerials. for both position and polarization measurements.

In these existing interferometers the position of the source is calculated from the interference pattern, which is produced by motion of the source through the lobe pattern of the aerials. But if, as in the case of bursts, the source has a life of only a few seconds, it will not have moved sufficiently during this time to produce the required interference pattern. Consequently, these methods are not directly applicable where the disturbance is of short duration or is rapidly changing its position, and this paper will describe a modified spaced-aerial interferometer which has been built to measure the position and polarization of such a variable component.*

Two similar horizontally-spaced aerials are used, and the interference pattern is obtained by sweeping the lobe pattern of the system across the source 25 times per second. This is achieved by continuously changing the phase of one signal, relative to the other, through $1\frac{3}{4}$ wavelengths, giving a total shift of $1\frac{3}{4}$ lobes in the lobe pattern. The use of two different aerial spacings and a special calibrating technique allows the source to be located to within about 2 minutes of arc. The polarization of the radiation can be measured with the equipment, and a complete record of position and polarization for each burst is obtained in 1 second.

The equipment can also be used as a fixed-lobe interferometer if so desired, and in this arrangement the angular size of the source can be obtained.

Using this equipment, observations have been made during the period between May 1949 and August 1950. The detailed results of these will be given in separate papers of this series (Payne-Scott and Little 1951).

II. INTERFEROMETER PRINCIPLES

(*a*) *Fixed-lobe Interferometer*

The mode of operation of a two-aerial interferometer can be studied with the aid of a diagram such as Figure 1, in which A_1 and A_2 are two similar aerials, placed along a horizontal east-west line. Then if a point source, say on the Sun, is at an angle θ to this line, the path difference φ, between two rays R_1 and R_2 assumed parallel, is given by

$$\varphi=\frac{d}{\lambda}\cos\theta \text{ wavelengths,} \qquad \ldots\ldots\ldots\ldots\ldots (1)$$

---

* A brief description of this equipment was given at the U.R.S.I. conference at Zurich in September 1950.

where d = separation between aerials,

λ = wavelength.

Suppose the aerials are connected by electrically equal lengths of cable C_1 and C_2 to a receiver D. When φ is an odd number of half-wavelengths, the receiver output is a minimum; similarly, when φ is an even number of half-wavelengths, it will be a maximum. As the source moves across the sky, the receiver output will undergo a sinusoidal variation and we may regard the source as moving through a lobe pattern of the type shown in polar form in Figure 2 (a). Then the angular separation, $\Delta\theta$, between the central lobes is given approximately by the formula

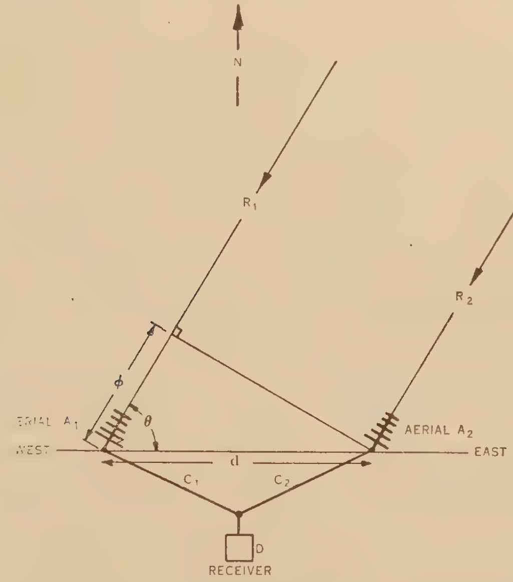


Fig. 1.—The two-aerial interferometer, showing the path difference, φ, between the two rays R_1 and R_2.

$$\Delta\theta = \lambda/d. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

It is interesting to note that if the beam width of an aerial system is defined to be the total angular width of the central lobe, then for the same beam width an interferometer requires a spacing equal to half the length of a broadside array (Fig. 2 (b)). From this figure it can also be seen that an interference system, even with quite small aerials, allows high directivity to be obtained. However, this is obtained at the expense of a possible ambiguity in position since with large numbers of lobes it becomes difficult to determine on which lobe the source lies. Thus, when $\Delta\theta$ is of the order of one Sun diameter or less, there is more than one possible value for the position of the source, i.e. there is more than one lobe on the Sun. For values of $\Delta\theta$ much larger than this, the Sun is small compared to a lobe and it becomes difficult to locate accurately small sources on the Sun's surface.

In the present system, the accuracy is determined mainly by the difficulty of reading the positions of the minima from the record, and is about 5 per cent. This is independent of the lobe spacing, and thus for an accuracy of 2 minutes of arc in measuring the position of a source, $\Delta\theta$ has to be about 40 minutes of arc. It should be noted that for this spacing, there will not be a marked interference pattern if the Sun's brightness is distributed uniformly over the disk, which is sometimes an advantage when observing point sources on the Sun.

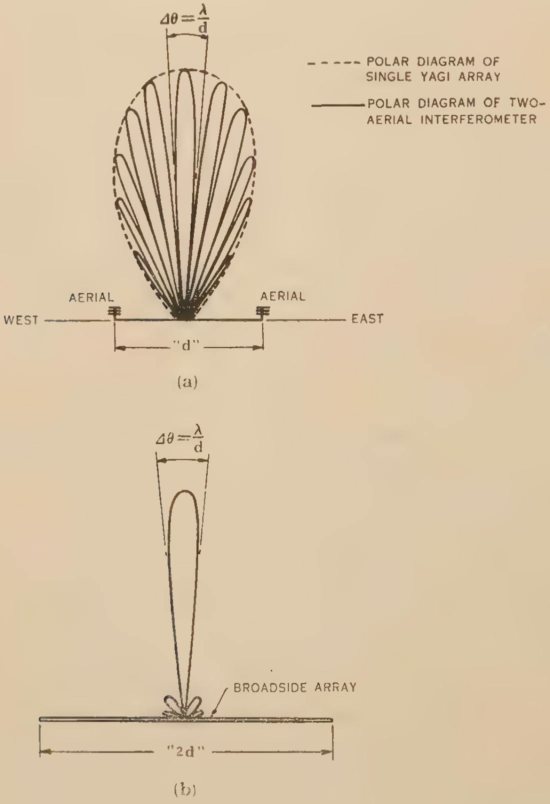

$\Delta\theta=\frac{\lambda}{d}$
POLAR DIAGRAM OF SINGLE YAGI ARRAY
POLAR DIAGRAM OF TWO-AERIAL INTERFEROMETER
AERIAL
AERIAL
WEST
EAST
"d"
(a)
$\Delta\theta=\frac{\lambda}{d}$
BROADSIDE ARRAY
"2d"
(b)

Fig. 2.—Polar diagrams of an interferometer and a broadside array.

Having chosen this value for our system, we find that there is a possible ambiguity in locating the source. To resolve this, a second aerial spacing is used, giving a different set of lobes, some of which will coincide with lobes in the original system. If the central lobes of the two patterns coincide, there will be other coincident lobes when

$$n\Delta\theta_1=m\Delta\theta_2, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

where n and m are integers. We can assume that the source is not more than a few radii off the solar disk, and hence the source is located uniquely from one pair of these coincident lobes as long as $\Delta\theta_1$ and $\Delta\theta_2$ are chosen so that coincident lobes are further apart than a few solar radii. This amounts to choosing $\Delta\theta_1$ and $\Delta\theta_2$ so that one is not an integral multiple of the other, nor can they be expressed as the ratio of two small integers. In practice the choice of $\Delta\theta_1$ and $\Delta\theta_2$ is influenced by the fact that coincidence is considered to occur when the lobes approach to within a small value set by the observational uncertainty, a, i.e.

$$| n\Delta\theta_1 - m\Delta\theta_2 | < a. \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

The receiver output when plotted as a function of time constitutes the interference pattern, and from it the hour-angle H of the source may be calculated, using the formula

$$\sin H = \frac{N\lambda}{d \cos \delta}, \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

where N is the number of the interference maximum from transit and δ is the declination of the source. Thus from the interference pattern for each spacing, a set of possible values of H is obtained, and the true position of the source can then be evaluated by comparing these two sets of results. If the declination of the source is not known, then the position is not a point but a curve which, if the source is assumed to lie within the limits of the Sun's declination, is for practical purposes a line parallel to the hour-circle through the centre of the solar disk.

In the present system the accuracy is 2 minutes of arc and six lobes of one spacing coincide, within the experimental limits, with seven of the other. With this arrangement it is found that when the two sets of possible positions are compared there are coincident values spaced about 4°, i.e. about 8 solar diameters.

(*b*) *Swept-lobe Interferometer*

So far the interference pattern has been considered as being produced by motion of the source through the lobe pattern of the aerials. But, as pointed out earlier, it is impossible to use this method for obtaining interference patterns from individual bursts. Referring back now to Figure 1, if the electrical length of cable C_1 is not equal to that of C_2, a phase difference is introduced into the system and the receiver output is no longer a maximum for $\theta=\pi/2$, but for some other value, θ_1. Further, if we vary the electrical length of, say C_1, continuously and at a uniform rate, the value of θ for which the receiver output is a maximum will be varied continuously. This has the effect of sweeping the lobe pattern across the source, and if the total phase change is one wavelength, it will be swept through one lobe. The interference pattern in this case has an insufficient number of fringes for easy recognition, and to facilitate measurements it is better to have a total phase change of more than one wavelength. Repeating this process rapidly enables interference patterns to be obtained from bursts.

III. THE INTERFEROMETER SYSTEM

(*a*) *The Site*

An interferometer operating on these principles has been set up on flat ground along the edge of one of Sydney's water-supply reservoirs, where the electrical noise is low. The aerials are on a surveyed east-west line and the receiving equipment is housed in a hut near the centre as shown in Figure 3.

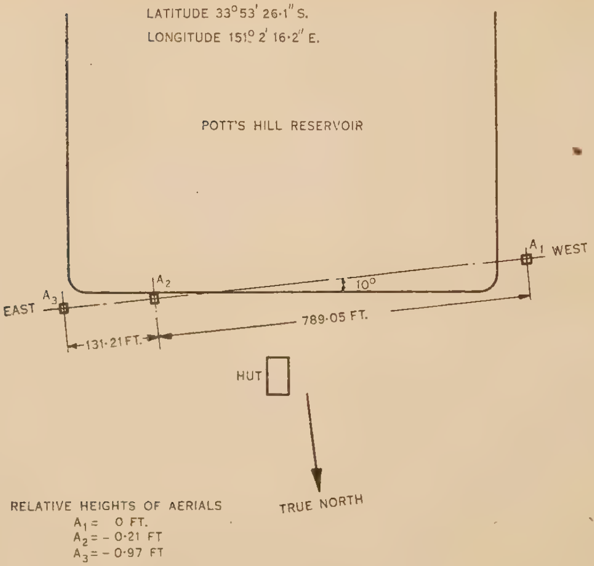

LATITUDE 33°53′ 26·1″ S.
LONGITUDE 151°2′ 16·2″ E.
POTT'S HILL RESERVOIR
A1 WEST
A2
A3
EAST
10°
789·05 FT.
131·21 FT.
HUT
TRUE NORTH
RELATIVE HEIGHTS OF AERIALS
A1 = 0 FT.
A2 = − 0·21 FT
A3 = − 0·97 FT

Fig. 3.—Disposition of the interferometer aerials.

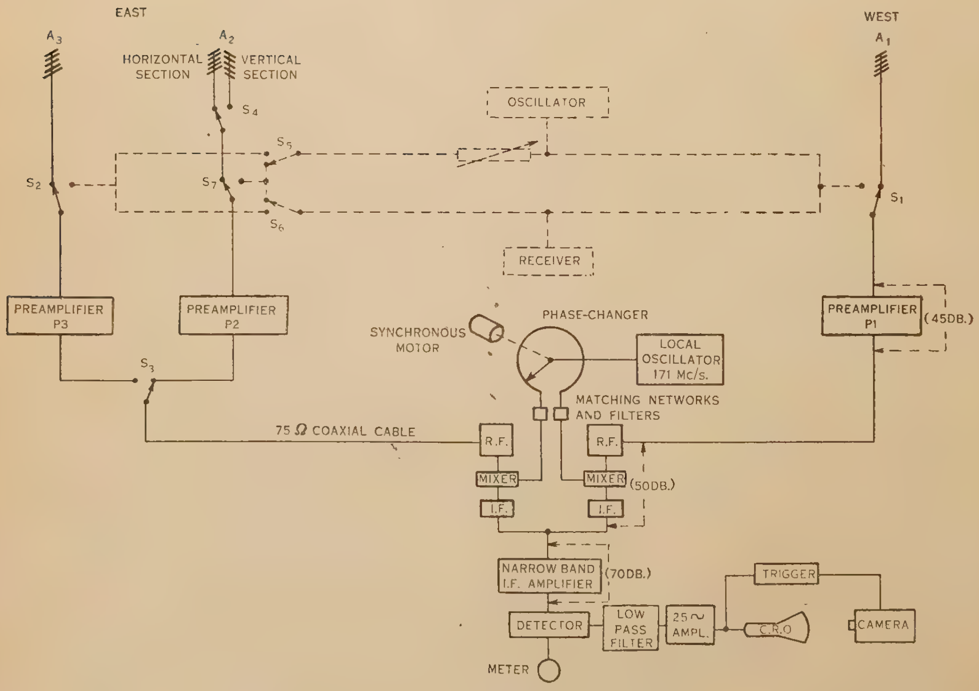

EAST
WEST
A3
A2
A1
HORIZONTAL SECTION
VERTICAL SECTION
S4
S5
S7
S6
S2
S1
S3
OSCILLATOR
RECEIVER
PREAMPLIFIER P3
PREAMPLIFIER P2
PREAMPLIFIER P1
(45DB.)
SYNCHRONOUS MOTOR
PHASE-CHANGER
LOCAL OSCILLATOR 171 Mc/s.
MATCHING NETWORKS AND FILTERS
75 Ω COAXIAL CABLE
R.F.
R.F.
MIXER
MIXER
(50DB.)
I.F.
I.F.
NARROW BAND I.F. AMPLIFIER
(70DB.)
TRIGGER
DETECTOR
LOW PASS FILTER
25 ∼ AMPL.
C.R.O
CAMERA
METER

Fig. 4.—Schematic of swept-lobe interferometer.

Three aerial spacings are available, but for the present purposes only the two spacings (A_1 to A_2) and (A_1 to A_3) are used.

(*b*) *The Swept-lobe Interferometer*

The swept-lobe interferometer equipment is shown in the block diagram of Figure 4. The operating frequency was chosen to be 97 Mc/s., the choice being restricted slightly by the proximity of a transmitter.

The aerials, A_1, A_2, A_3, consist (Plate 1) of five-element crossed Yagi arrays, which are motor-driven to follow the Sun. Normally only the horizontal section is used, but when polarization observations are being made one of the aerials (A_2) is switched to the vertical section ; it was thought desirable to have all the aerials physically identical in order to minimize any phase differences between them. They can be connected in the combinations A_1 and A_2, or A_1 and A_3, by operating the coaxial switch S_3, whilst the switch S_4 serves to change the aerial A_2 from horizontal to vertical polarization.

Each aerial has its own preamplifier consisting of a three-stage " cascode " amplifier of the type described by Wallman, MacNee, and Gadsden (1948) and having a gain of 45 db. with a noise factor of 4·0 db. The outputs from these are fed into equal lengths of 75 ohm polythene coaxial cable, type " Uniradio No. 1 ", having a loss of about 12 db. The signals are equalized by varying the voltages applied to the anodes of the preamplifier valves.

As mentioned earlier, interference patterns from bursts can be obtained by varying the phase of one signal relative to the other, and it can be shown that for two receivers having a common local oscillator, the same effect is produced by varying the relative phases of the two local oscillator signals. Thus, when the two intermediate-frequency signals are connected to a common terminal, the resultant voltage is given by an expression of the form

$$E_{(\mathrm{I.F.})}=A\sin(\omega_2-\omega_1)t\cos\frac{(\alpha_1-\alpha_2)-\Phi}{2}, \quad \ldots\ldots\ldots\ldots (6)$$

where $\omega_2-\omega_1=\omega_{(\mathrm{I.F.})}$, the intermediate frequency, Φ is the phase difference between the two signals of frequency ω_1, $(\alpha_1-\alpha_2)$ is the phase difference between the two local-oscillator signals of frequency ω_2, and $(\alpha_1+\alpha_2)$ is the total phase length of the phase-changer.

This method has proved to be satisfactory for the problem and the actual phase-changer used consists (Plate 2) of a 300 ohm balanced line, about 7/8 of a wavelength long, and bent to form a circle. The line is connected between the mixers of two receivers, and the local oscillator is capacitively coupled to the line by means of two small plates. These are mounted on an insulated arm and connected to a rotating joint formed by two concentric loops. The arm is driven by a synchronous motor at 1500 r.p.m. and thus the coupling point is continuously varied, giving a total relative phase-change of twice the electrical length of the line, 25 times per second. The advantage of phase-changing in the local-oscillator circuit is that amplitude variations in the phase-changer will have only a small effect on the interference pattern. Again, for a local oscillator of higher frequency than the signal, this type of phase-changer is physically smaller than one at the signal frequency, which is a desirable feature for mechanical reasons.

The choice of the bandwidth of this system is restricted and lies between two well-defined limits. The lower limit to the overall frequency response is about 50 c/s. and is fixed by the necessity to display the 25 c/s. interference pattern without distortion.

The widest bandwidth allowable is set by the fact that solar noise is not monochromatic. The central interference maximum is the same for all frequencies but the spacing between maxima is frequency dependent. Consequently the maxima and minima are blurred at times when the Sun is distant from the meridian. Moreover, if the receiver passband at such times is asymmetrical there will be also a shift in the maxima and minima. This shift is equivalent to a change Δf in the centre frequency and from equation (5) the corresponding error in the position of the source is

$$\Delta\beta=\frac{\Delta f}{f}\tan H. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (7)$$

The change Δf cannot be greater than the semi-bandwidth and is usually considerably less. For the worst possible case, if the position error is not to exceed $\Delta\beta$ for any hour-angle between H and $-H$ then the receiver bandwidth must not exceed twice the value of Δf given by equation (7). In particular, if observations are to be made for 2 hours before or after transit with an accuracy of 1 minute of arc, the maximum allowable bandwidth is 200 kc/s. We have used a bandwidth of 150 kc/s.

It should be pointed out that although, in general, solar noise is not monochromatic, Wild (1951) has shown that bursts do have a finite bandwidth sometimes as small as a few megacycles per second. Hence the bandwidth of the equipment should be small compared to this if confusion between bursts on neighbouring frequencies is to be avoided. The 150 kc/s. used satisfies this condition.

The aerials are connected through the preamplifiers and cables to two identical receivers with a common 171 Mc/s. local oscillator coupled through the phase-changer (Fig. 4). Filters are provided to prevent signals from one receiver being coupled through the phase-changer into the other. The 74 Mc/s. intermediate frequency outputs of the receivers are mixed and fed into an amplifier, having a bandpass of 150 kc/s. After detection the output is integrated and coupled through a two-stage low-frequency amplifier to the vertical deflection plates of a cathode-ray tube, the horizontal deflection for which is a 25 c/s. linear time base, synchronized to the mains. The gains of the respective amplifiers are indicated on Figure 4, on the side connected to aerial A_1.

Thus, for a source of small angular size, rotation of the phase-changer produces a sine wave on the cathode-ray screen. The position of the minimum will vary with the position of the source and for record purposes the screen is photographed, together with a clock, a film frame counter, and a meter registering receiver-output current. The camera is a modified 16 mm. cine-camera, which is triggered automatically whenever the intensity of radiation rises above a predetermined value, and is arranged to take three consecutive photographs for each trigger pulse. The first is the interference pattern from aerials A_1 (horizontal) and A_2 (horizontal). After completing this, aerial A_2 is automatically switched to its vertical section and the camera takes a photograph with A_1 (horizontal) and A_2 (vertical). Having done this, A_2 is replaced by A_3 and a photograph is taken of the interference pattern from A_1 (horizontal) and A_3 (horizontal). The total time taken to complete the sequence is about one second. The position of sources with intensities less than about 5×10^{-21} W.m.$^{-2}$ (c/s.)$^{-1}$ cannot be measured with this equipment since there are small amplitude fluctuations as the phase-changer rotates and these cause variations in the receiver output. These are of comparable size to the interference pattern and the apparatus is thus inaccurate for such low intensities. At present when we require rapid measurements the intensity of radiation is usually high, while for continuous radiation of low intensity the fixed-lobe interferometer can be used. However, if the method is to be used for rapid measurements of sources of very low intensity a better phase-changer than we have at present would be necessary.

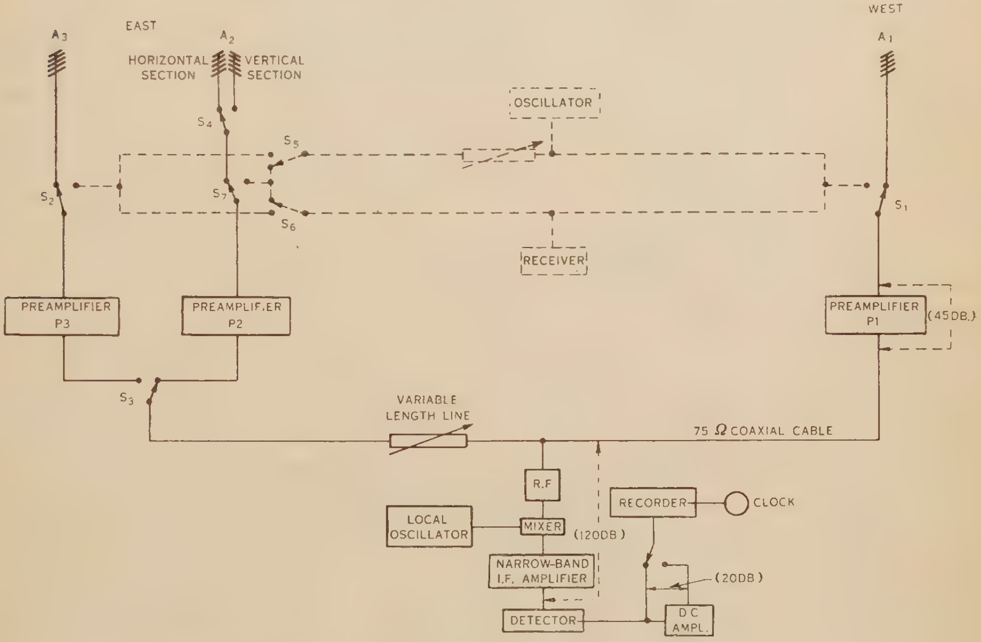

WEST
EAST
A_3
A_2
A_1
HORIZONTAL SECTION
VERTICAL SECTION
OSCILLATOR
S_4
S_5
S_2
S_7
S_6
S_1
RECEIVER
PREAMPLIFIER P3
PREAMPLIF.ER P2
PREAMPLIFIER P1
(45DB.)
S_3
VARIABLE LENGTH LINE
75 Ω COAXIAL CABLE
R.F
LOCAL OSCILLATOR
MIXER
(120DB)
RECORDER
CLOCK
NARROW-BAND I.F. AMPLIFIER
(20DB)
DETECTOR
DC AMPL.

Fig. 5.—Schematic of fixed-lobe interferometer.

(c) *Fixed-lobe Interferometer*

Sometimes when there is a steady level of radiation lasting for more than a few minutes, the equipment is used as a normal fixed-lobe interferometer. Here (see Fig. 5) the cables from the preamplifiers are connected to a common 97 Mc/s. receiver having a bandpass of 150 kc/s. The output of this receiver can be connected directly to a 0–1 mA. Esterline-Angus recorder, or can be connected through a direct current amplifier to the recorder if the intensity is very low. Time marks are placed on the record from a mechanical clock, checked against standard time signals.

The dotted section of Figures 4 and 5 refers to the calibration system, which will be described in the next section.

IV. CALIBRATION OF THE INTERFEROMETER

In a system such as that described, the phase difference between the eastern and western halves must be known before any accurate measurements can be made. Furthermore, this phase difference may not stay constant, owing to changes in the preamplifiers and cables—the change in the latter case being due mainly to the temperature coefficient of the cables. With our cables, which are not buried, this is the major cause of drifts. Measurements have shown that the rate of change of electrical length can be as much as that equivalent to a shift of the source of 2 minutes of arc per hour during the middle of the day, the period during which the equipment is used. It becomes necessary, therefore, to calibrate the cables about once every half-hour. In the future, however, we shall use a higher quality cable which will be buried, and which, as tests have indicated, should not require calibration more than about once every two days.

(a) *Calibration Method*

It would be possible to calibrate by radiating a signal from an aerial at a known point in space, but this aerial would have to be at so great a height and distance to eliminate effects due to different ground reflections at each aerial that the method would be very difficult to use in practice. However, a method has been developed by which it is possible to calibrate the whole system quickly, with the exception of the aerials. These are calibrated separately by a method which will be described later.

The calibration of the system is carried out using the arrangement shown in Figure 6, in which E and F are the two signal paths, including preamplifiers, and A, B, C, and D are 75 ohm calibrating cables, each having a loss of about 13 db. P and L are variable-length coaxial lines of constant 75 ohm impedance. The coaxial switches connect the preamplifiers either to the aerials or to the calibrating cables.

The problem is to evaluate the phase difference between the two signal paths E and F. In the swept-lobe interferometer we have to determine at what point along the photographed interference pattern, signals arriving at the two preamplifiers in the same phase would produce a minimum, and in the fixed-lobe interferometer it is necessary to make the two signal paths equal by adjusting the length of L. This is achieved by using signals sent along the calibrating cables.

Consider the cables A, B, C, and D; with the preamplifiers switched to the aerials, there are short circuits at R and S which are transformed by quarter wavelength, 53 ohm cables to open circuits at X and Y. The line-lengthener P is adjusted so that the phase-receiver shows a minimum (in practice it is easier to set to a minimum than to a maximum) and we then have, using the letters to denote electrical lengths,

$$A+C=B+D\pm(2N+1)\lambda/2. \qquad (8)$$

The odd number of half-wavelengths, $(2N+1)\lambda/2$, arises from setting to a minimum and not a maximum.

The preamplifiers are next switched to the calibrating cables through the $\lambda/4$, 53 ohm transformers which match the 75 ohm preamplifiers to the 37·5 ohm junctions X and Y. A signal is sent round the two paths A–E and B–F, and then, interchanging the oscillator and the phase-receiver, round the two paths C–E and D–F; in the case of the swept-lobe interferometer a photograph is taken at each stage and the position of the minima measured. A little consideration will show that, because of the relationship between the alternative paths, the mean of these two minima will differ by $\pm\lambda/4$ from the minimum which will be obtained from a signal reaching each preamplifier in the same phase—i.e. we have succeeded in finding the point of zero phase difference for the system.

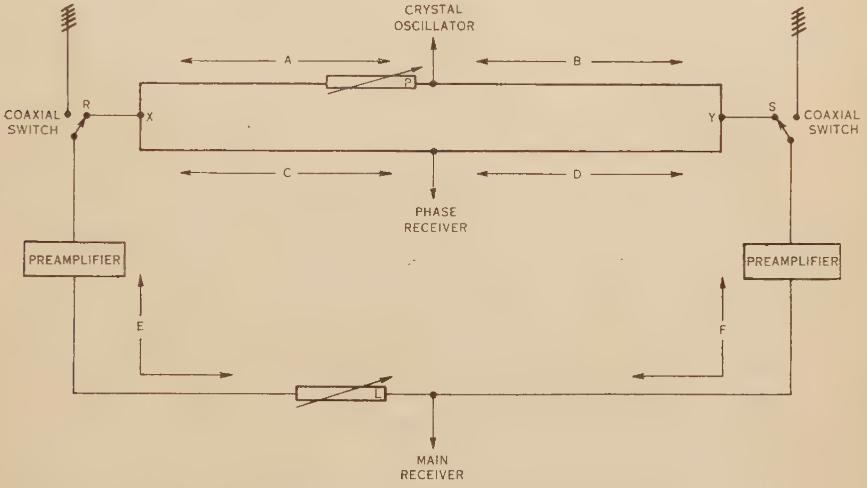


Fig. 6.—Simplified diagram of calibration system.

In the case of the fixed-lobe interferometer, the length of L is adjusted to the values L_1 and L_2 in turn to give a minimum for each of the alternative paths in the calibrating process, and is finally set at the value $\frac{1}{2}(L_1+L_2)$, which is thus $\pm\lambda/4$ from the true zero. The ambiguity in sign is inherent in this method of calibrating, in which all signals are returned for measuring to the place at which they were generated after travelling to and from the preamplifiers, but this feature of generation and measurement at the same place also makes the method very easy to operate.

The application of the above method depends on the fact that the cables have a high loss and may therefore be considered as infinite lines. If this were not the case, the above expressions would need to be modified to include waves travelling in the opposite directions, arising from reflections.

The complete calibrating system is shown in dotted outline in Figures 4 and 5; the switches S_5 and S_6 are used to connect the calibrating cables to either of the eastern preamplifiers. The short lengths of cables in the aerial side of the switches are made electrically equal and are assumed to remain constant.

(*b*) *The Coaxial Switches*

The switches used are not of the elementary form indicated in Figures 4, 5, and 6, since it is difficult to construct this type without introducing serious mismatches, but are of the form shown in Figure 7. Each switch consists of two short sections of 75 ohm coaxial line, either of which can be short-circuited at its centre by means of a solenoid-operated plunger, thus allowing the switches to be remote-controlled.

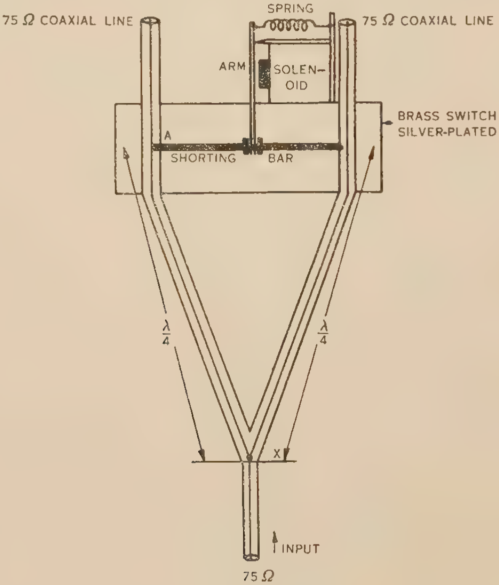


Fig. 7.—Details of coaxial switch.

In operation, when the cable at A is short-circuited, signals are blocked from travelling along the line and at the same time the shunt admittance of this cable at X is negligible. Under these conditions signals can travel unimpeded along the other line. The voltage standing-wave ratio of these switches is better than 1·1 to 1 and consequently they do not introduce any appreciable discontinuities into the system.

(*c*) *Aerial Calibration*

When the system has been calibrated up to the aerial switches, only the aerials and their short connecting cables remain uncalibrated. Now these are matched and made identical, as far as is practicable, but a direct check is made to determine any phase differences between them.

One of the aerials is turned so as to lie in a horizontal plane and point due north, as in Figure 8. At a distance of about 15 ft. another Yagi of the same polarization and height above ground is set up, and connected through an attenuating pad to a crystal-controlled oscillator. The received signal from the aerial under test is passed through another pad to a line *M* of variable length, the other end of which is connected to the oscillator. The signals at *A* and *B* are adjusted to be approximately equal by the equalizing attenuator, and then connected through the pad to the receiver. " *M* " is now adjusted so that the received signal is a minimum, and the position of *M* is noted. The whole apparatus is moved to each of the other aerials in turn and the measurement repeated for the same conditions.

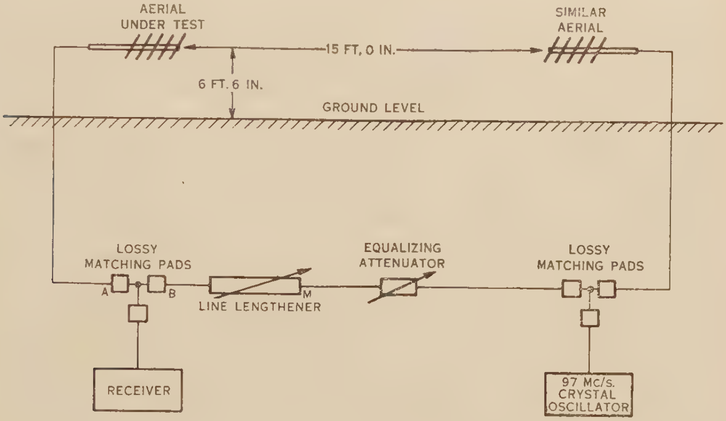


Fig. 8.—Method of aerial calibration.

The results of the experiments which were carried out showed that there is no appreciable difference between the aerials used. It is not considered necessary to repeat this calibration more often than about once every 12 months.

V. Observational Technique

Observations have been made with this equipment over the period between May 1949 and August 1950. During this time the interferometer has been operated every week-day for about 2 hours on either side of meridian transit and in periods of high activity, observations have also been made at week-ends. Records have been taken of 30 noise storms, 6 outbursts, and 25 randomly-polarized bursts.* In the case of the noise storms, the number of records exceeds the number of storms since these latter often last for several days. About 50 per cent. of the 25 randomly-polarized bursts occurred during noise storms whilst the remainder occurred in the absence of any other activity (as determined by the sensitivity of our equipment). Similarly, one of the outbursts occurred at the beginning of a noise storm and another at the end of a storm whilst the other four occurred without any other activity being visible. From all these records, the position and polarization of the sources have been calculated, and for a few of the noise storms, a value for the physical size was also obtained.

---

* This type of burst has been referred to by Pawsey (1950) as an isolated burst, whereas Payne-Scott (1949) has called them unpolarized bursts, meaning not circularly polarized.

(*a*) *Position Measurements*

Before any observations are made, the signals from the preamplifiers are equalized and the equipment is calibrated. As described in the previous section, this amounts, in the case of the swept-lobe interferometer, to photographing two calibrating interference patterns. The mean of the two minima or of L_1 and L_2 is the reference or mid-point of the system, from which all other minima are measured ; there is one reference point for each aerial spacing. Typical calibration patterns are reproduced in Plate 3, which also shows the position of the reference point for this particular calibration. After calibration, the sensitivity is set to the required operating value and the camera set to trigger whenever the level reaches some predetermined value. During noise storms a steady enhanced level often appears which would cause the camera to take pictures continuously. In this case it can be arranged to trigger every minute, whilst still being able to trigger from bursts which may occur above this level of radiation.

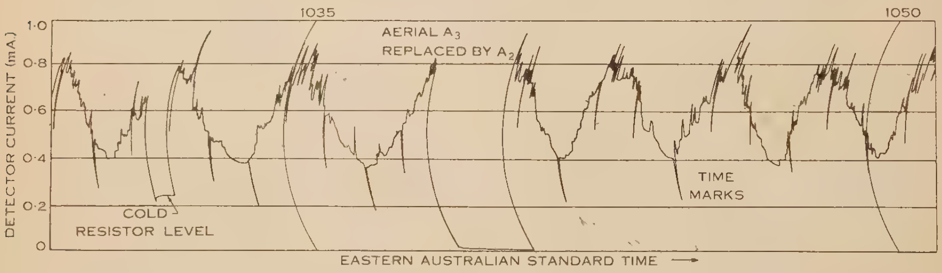


Fig. 9.—Interference pattern taken with the fixed-lobe interferometer.

Sometimes, as mentioned earlier, the equipment is used as a normal fixed-lobe interferometer. Here again it is necessary to calibrate before recording, and using the calibration procedure outlined in the previous section, the value of $(L_1+L_2)/2$ is found for one pair of aerials and L is set at this value. The equipment is allowed to run until about six minima are produced. The value of $(L_1+L_2)/2$ for the other pair of aerials is then found, and after resetting L to this new value, six minima for this spacing are recorded. Alternation of the aerial spacings is thus carried out during the whole observing time. A typical record is shown in Figure 9. The downward kicks are time-marks at minute intervals.

(*b*) *Polarization*

The method used was first introduced by Ryle and Vonberg (1948) for circular polarization measurements, but has been extended here to include other polarizations. The change in the interference pattern is observed when one of the two horizontally polarized aerials, A_2, is replaced by a vertically polarized one (A_1 horizontal and A_2 vertical).

If the radiation is circularly polarized, the interference pattern will shift through $\lambda/4$, or a quarter of a lobe, without any change in amplitude. The direction of the shift determines whether the radiation is right- or left-handed circularly polarized. An example is shown in Plate 4* which is taken from the record of a noise storm.

For randomly polarized radiation, the pattern disappears completely, since there is no phase coherence between the signals received by each aerial. At the same time there is no change in the received intensity, and consequently no change in receiver output. This is shown clearly in Plate 5, which is a record of an " unpolarized " burst.

For a linearly polarized wave, the phase-change in the interference pattern is either zero or half a lobe, while there will usually be a change in amplitude, the amount depending on the inclination between the plane of polarization and the aerial planes. Plate 6* shows an example of this taken from the record of a noise storm.

Any combination of random and elliptical polarization (including in elliptical the particular cases of circular and linear) can be elucidated by comparing on the two photographs taken with parallel and crossed aerials the three measurable quantities—relative phase of interference patterns, relative amplitudes of interference patterns, and relative amplitudes of detector current (read by meter). For example, in the particular case of elliptical polarization in which the direction of the axes of the ellipse coincides with those of the aerials, the phase displacement is a quarter of a lobe. The same displacement occurs for circular polarization or a mixture of circular and random, but in the latter case the detector current remains unchanged, while in the case of elliptical polarization it alters.

(c) *Measurement of Source Size*

An effect resembling that of a finite bandwidth is produced by an extended source, which may occupy a large fraction of a lobe. The interference pattern from such a source is made up of the sum of separate patterns from each point on the source, and since each of these is displaced from its neighbour, the resultant pattern has maxima and minima which are not sharply defined. The ratio of maximum to minimum is reduced and is therefore a measure of the source size. Actually it is a measure only of the width of a vertical strip, containing the source.

To calculate the width requires a knowledge of the relative power of the maxima and minima, and therefore the receiver-output current for each must be known, as well as the cosmic- and receiver-noise levels.

For sources which are small compared to the distance between lobes, the deviation from complete cancellation at the minima is small and must be measured very accurately. Such a deviation can be produced not only by an extended source, but by small differences in the noise factors of the two preamplifiers. Also the length of the interference pattern is not an integral number of wavelengths, and hence the receiver output measured by the meter varies slightly with the phase of the pattern. In view of these difficulties we have not so far attempted to measure source size with the swept-lobe interferometer, and at present all we can say about the size of a disturbance is that it is less than or comparable to the separation between lobes.

---

* For convenience, in Plates 4 and 6 the distances X and Y have been referred to one end of the trace and not to the reference line shown in Plate 3.

On some occasions source sizes have been measured using the fixed-lobe interferometer, the same method being used as previously by Pawsey, Payne-Scott, and McCready (1946), Ryle and Vonberg (1946), and McCready, Pawsey, and Payne-Scott (1947). There will be some error in these due to the slight difference in noise factor of the two preamplifiers, but this error becomes negligible for sources of high intensity.

(*d*) *Discussion of Errors*

The accuracy with which positions can be obtained depends on several factors, the most important of which can be summarized thus :

1. The accuracy with which the position of the maxima and minima can be measured from the record.
2. Changes in the zero of the system.
3. The errors away from transit due to :
    - (*a*) Uncertainty in declination of the source.
    - (*b*) Errors in measurement of frequency.
    - (*c*) Errors in measurement of spacing between the aerials.
4. Atmospheric and ionospheric refraction.
5. Ground reflections.

These errors are discussed in turn below.

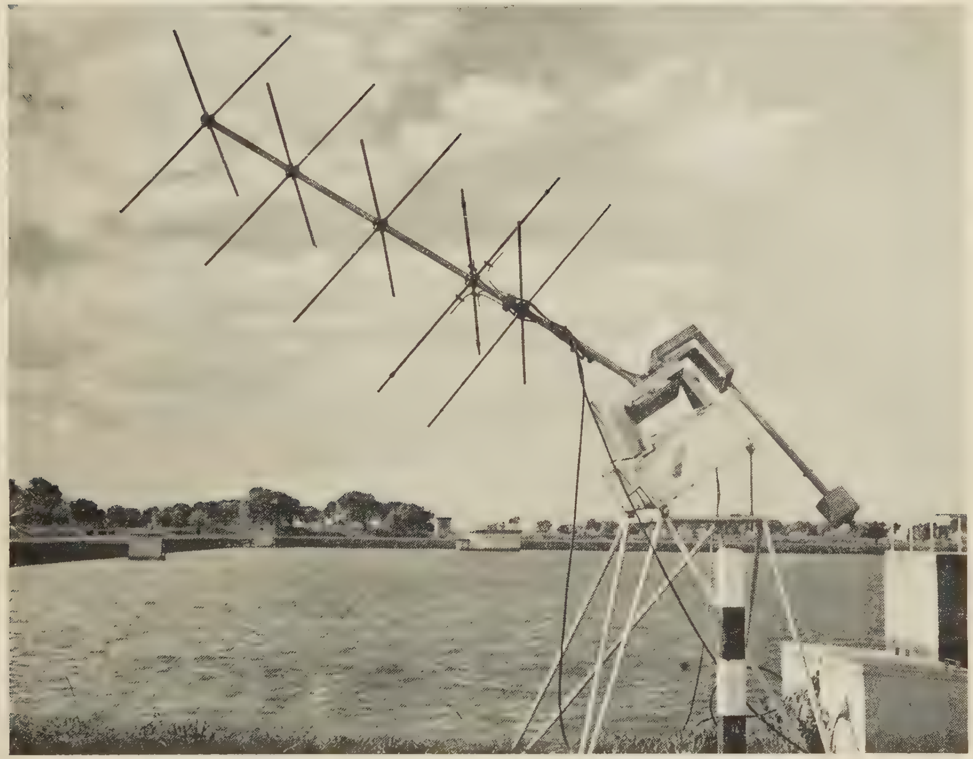
(1) To test the measurement accuracy, an artificial signal whose phase could be varied by known increments was used. The measured phase-changes agreed with the actual ones with a discrepancy corresponding to a position change of not more than 1 minute of arc, provided that the input signal corresponded to a source of at least 5×10^{-21} W.m.$^{-2}$ (c/s.)$^{-1}$ intensity ; for smaller sources irregularities in the phase-changer increase the error. The measuring error is due partly to lack of uniformity in the phase-changer and partly to the actual measurement uncertainty. The latter is reduced when a number of observations are made in rapid succession on one radiating source.

(2) As discussed in the section on calibration, the relative changes in zero are kept below 1 minute of arc by frequent recalibration. The observations of source position which will be described in the succeeding two parts of this series show no systematic position displacement in one direction from the optical sources and it is concluded that the absolute position of the zero is correct to 1 or 2 minutes of arc.

(3) From the formula

$$\sin H=\frac{N\lambda}{d\cos\delta} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (9)$$

MOVEMENT OF SOURCES OF RADIATION ACROSS THE SUN'S DISK. I



Western Yagi aerial.

MOVEMENT OF SOURCES OF RADIATION ACROSS THE SUN'S DISK. I

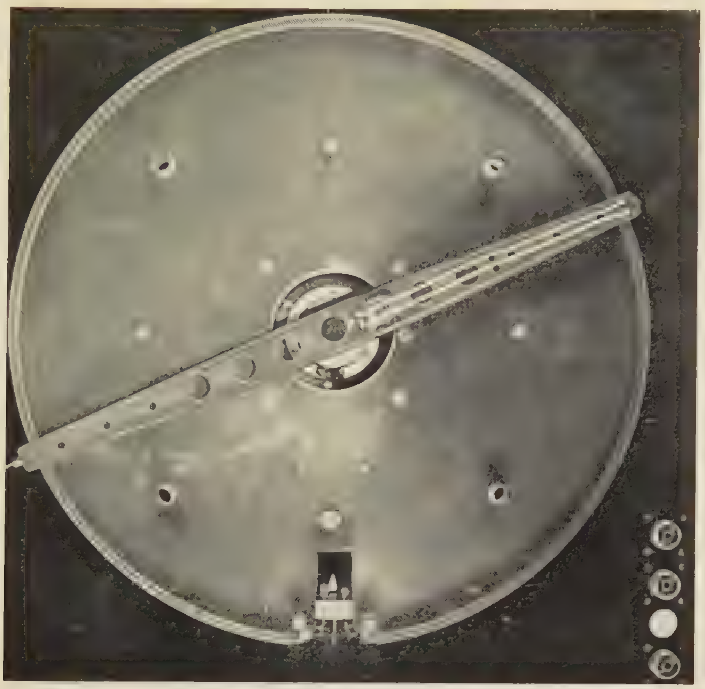


Fig. 1.—Phase-changer, showing inner coupling loop.

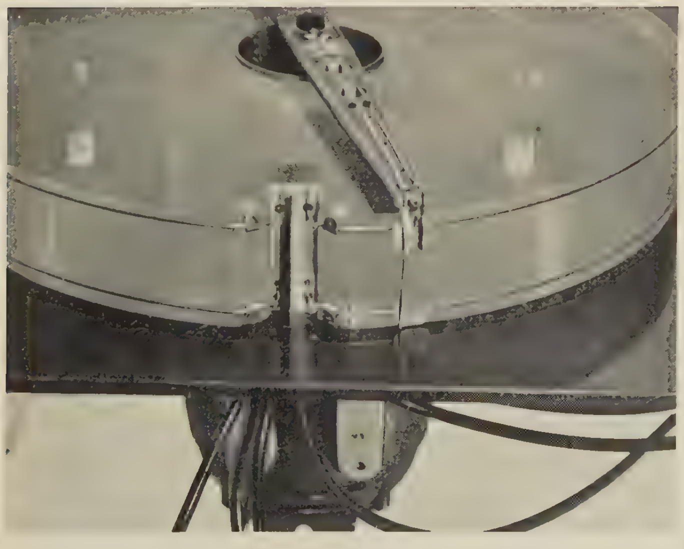
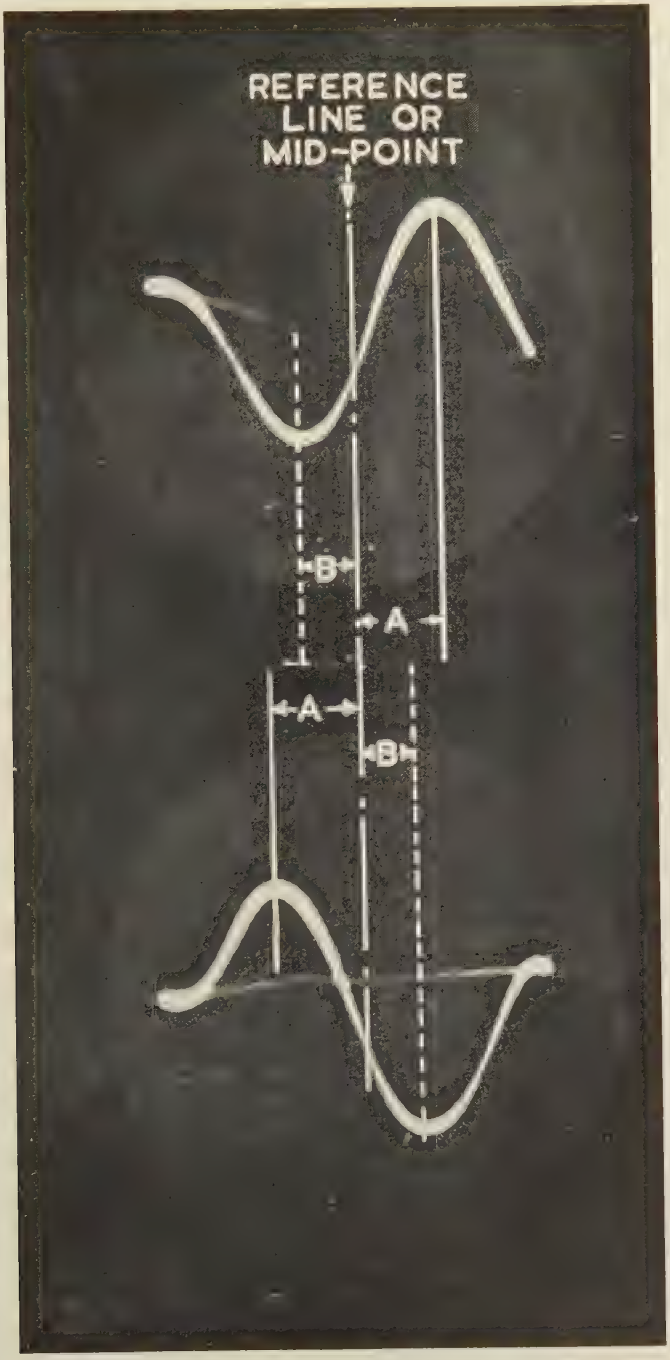


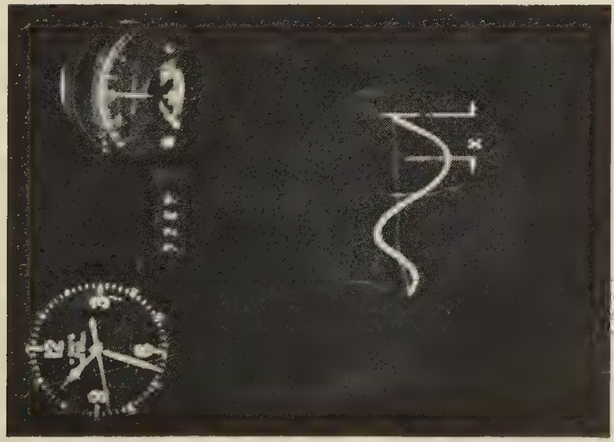
Fig. 2.—Phase-changer, showing capacity coupling to transmission line, and connections to ends of line.

MOVEMENT OF SOURCES OF RADIATION ACROSS THE SUN'S DISK. I

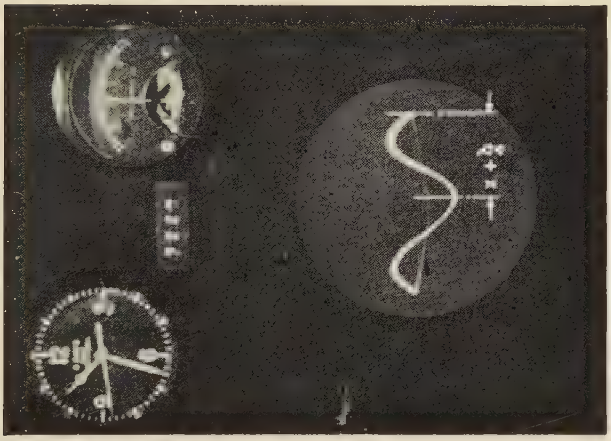

REFERENCE LINE OR MID-POINT
B
A
A
B

Typical calibration patterns.

MOVEMENT OF SOURCES OF RADIATION ACROSS THE SUN'S DISK. I

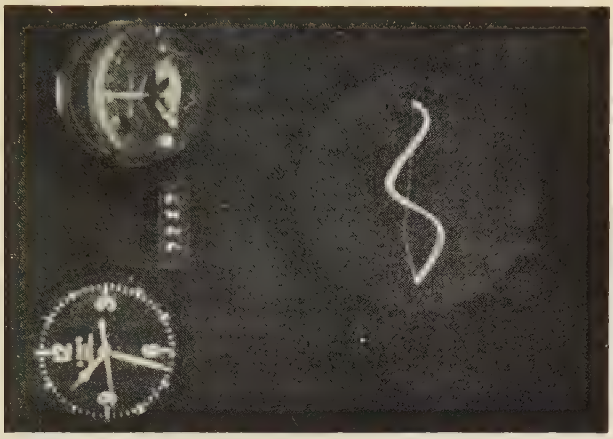


(a) AERIAL A_1-HORIZONTAL
AERIAL A_2-HORIZONTAL



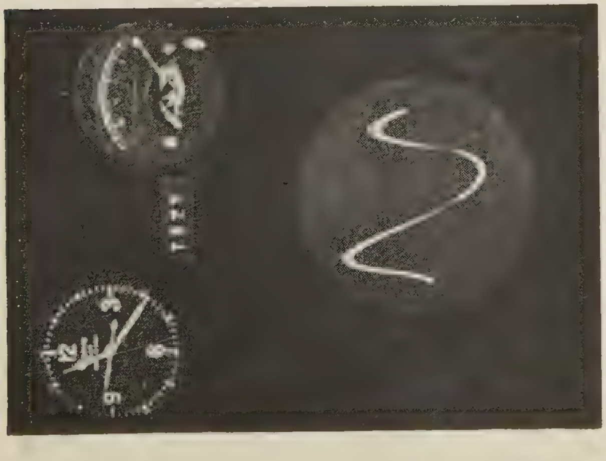
(b) AERIAL A_1-HORIZONTAL
AERIAL A_2-VERTICAL

Storm radiation—circularly polarized.

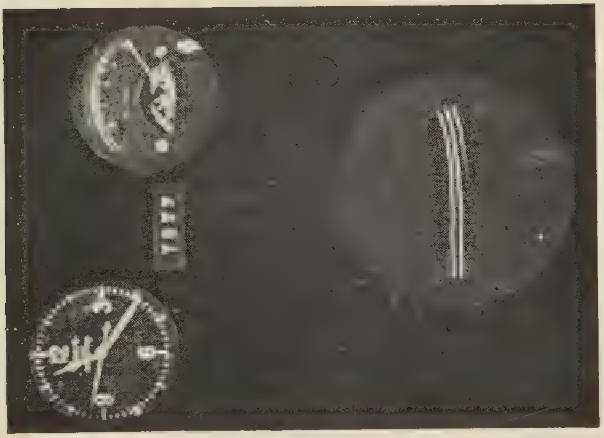


(c) AERIAL A_1-HORIZONTAL
AERIAL A_3-HORIZONTAL

MOVEMENT OF SOURCES OF RADIATION ACROSS THE SUN'S DISK. I

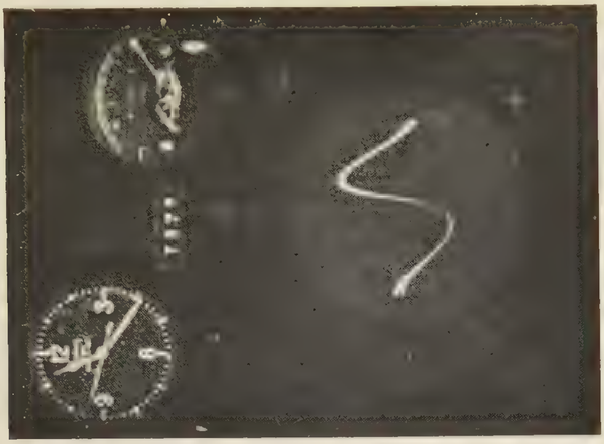


(a) AERIAL A_1–HORIZONTAL
AERIAL A_2–HORIZONTAL



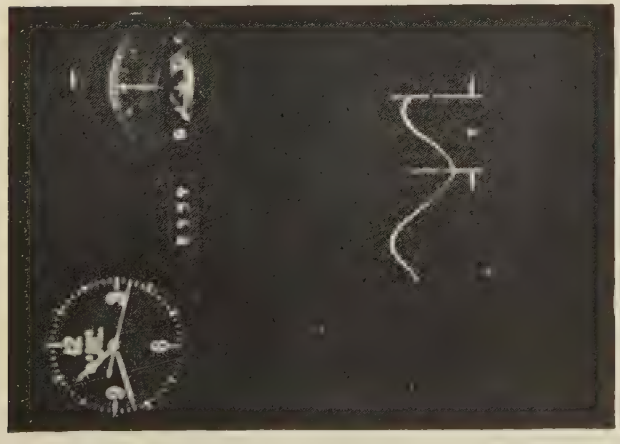
(b) AERIAL A_1–HORIZONTAL
AERIAL A_2–VERTICAL

A typical unpolarized burst.

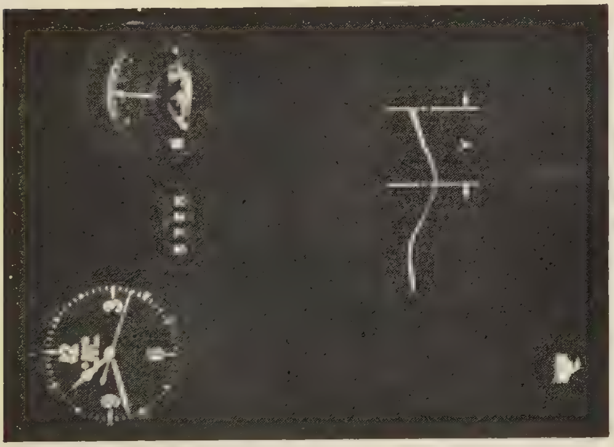


(c) AERIAL A_1–HORIZONTAL
AERIAL A_3–HORIZONTAL

MOVEMENT OF SOURCES OF RADIATION ACROSS THE SUN'S DISK. I

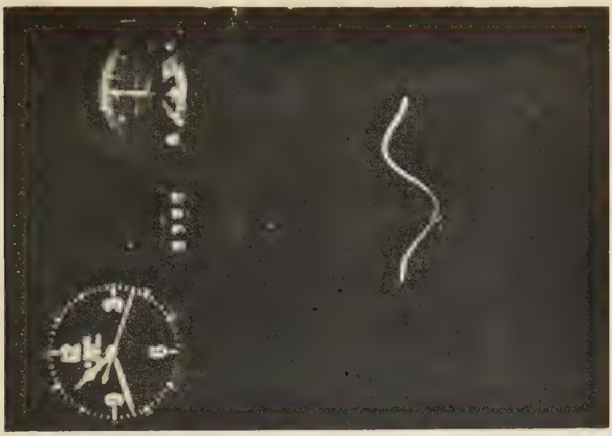


(a) AERIAL A_1-HORIZONTAL
AERIAL A_2-HORIZONTAL



(b) AERIAL A_1-HORIZONTAL
AERIAL A_2-VERTICAL

An example of linear polarization.



(c) AERIAL A_1-HORIZONTAL
AERIAL A_3-HORIZONTAL

it is apparent that as H increases, i.e. as the source moves away from transit, errors in λ, d, or δ produce continuously increasing errors in the observed position, determined by H. The equipment is not normally operated with H greater than 30° (i.e. 2 hours each side of transit).

(*a*) The declination of the source is taken as that of the centre of the Sun in first computing the position, but in practice the source is always close to a sunspot group and the results are finally corrected to correspond to a source with the declination of this group. In the worst case, when $\delta=\pm 23°$ and $H=\pm 30°$, an error of 1 minute of arc in declination produces an error of $\frac{1}{4}$ minute of arc in source position, while for $\delta=0$ the error becomes zero.

(*b*) The frequency is determined by a crystal controlling the frequency of the local oscillator, whose frequency is known to be $171\pm 0{\cdot}02$ Mc/s. and hence the maximum error in H from this source is $0{\cdot}4$ minute of arc. As explained earlier, the bandwidth is chosen so that errors due to asymmetry in bandwidth cannot be more than 1 minute of arc, and are, in fact, considerably less.

(*c*) The distance between the aerials is known to within $0{\cdot}25$ ft., allowing for some uncertainty in the position of the electrical centres of the aerials, and the resulting error in the location of the source is less than $\frac{1}{2}$ minute of arc.

(4) Refraction can be considered in two parts, that which would occur on a plane Earth and that due to the fact that the Earth is spherical. A uniform plane sheet of refractive material does not introduce any differential path difference between parallel rays, and hence the refraction due to a uniformly stratified region above a plane Earth can be ignored, provided the velocity of light *in vacuo* (i.e. outside the refracting region) is used in calculating wavelengths. We have taken this velocity as $2{\cdot}9979\times 10^{10}$ cm./sec.

The effect of the Earth's curvature remains to be considered. At our latitude the Sun at no season of the year has an elevation of less than 28° during the period of 2 hours on each side of transit. At such elevation the contribution of the Earth's curvature to the refraction caused by the lower atmosphere can be ignored, as can be seen by studying the formula given by McCready, Pawsey, and Payne-Scott (1947) or those given in any standard text on astronomy. On the other hand the component of ionospheric refraction due to the Earth's curvature is appreciable. Unfortunately the detailed structure of the ionosphere is not known but Bailey (1948), assuming a parabolic density of charge both above and below the maximum value for each layer, has calculated a formula for ionospheric refraction. Measurements by Payne-Scott and McCready (1948) on solar noise at low elevations indicate that Bailey's values are too low for low angles of elevation, but approach the observed values as the elevation increases. Since the lowest elevation at which our interferometer measurements are made is 28°, we have assumed that Bailey's formula is sufficiently correct. For such elevations this formula reduces to

$$\Delta E=\frac{2d}{3a}\left(\frac{f_0}{f}\right)^2\times\frac{1}{\sin^2\theta\tan\theta}, \qquad \ldots\ldots\ldots\ldots\ (10)$$

where ΔE is the change in the elevation E of the source due to refraction, d is the semi-thickness of the refracting layer and f_0 its critical frequency, a the

Earth's radius, and f the frequency of the source. The only appreciable refraction occurs in the F layer, and inserting

$$d=75 \text{ km.},$$

$$f_0=12 \text{ Mc/s.},$$

$$\Delta E=\frac{120\times10^{-6}}{\sin^2\theta\tan\theta}. \qquad (11)$$

From the fundamental formula for the interferometer it can be shown that the change ΔH in the hour-angle of a source of hour-angle H due to a change in elevation ΔE is

$$\frac{\Delta H}{H}=\Delta E\tan E, \qquad (12)$$

and hence

$$\frac{\Delta H}{H}=\frac{120\times10^{-6}\tan E}{\sin^2\theta\tan\theta}. \qquad (13)$$

Under our working conditions the maximum value of the error is 1 minute of arc, occurring in midwinter when the source is 2 hours before or after transit. Thus an error of even 100 per cent. in Bailey's formula will usually not be very serious.

(5) An error is introduced if ground reflected rays are accepted which are not identical for the two aerials. This error is negligible in midsummer, but in midwinter it becomes appreciable owing to the low elevation of the Sun and the fairly wide beam of the Yagi aerials. An estimate has been made of the consequent slight phase-changes in the received signal on the main pair of aerials (A_1 and A_2). This does not cause an error exceeding 2 minutes of arc and is usually not more than 1 minute of arc, as the ground near these aerials is fairly flat. However, for aerial A_3 there is a rising bank slightly east of north, and there are also power lines running north-south on the eastern side with the result that, in midwinter, deviations in position between the positions obtained on aerials A_1–A_2 and A_1–A_3 occur for about half an hour before local midday and are up to 3 minutes of arc. Since the effect of general reflections stays fairly constant over a week or so and is only large over a short period of time and for one aerial system, it can be practically eliminated when position measurements are made on a source of fairly constant position, or on a source radiating either continuously or at intervals for some days.

Finally, considering all these errors we conclude, that for sources of reasonable intensity, the position may be ascertained on a line on the Sun parallel to the hour-circle through the centre of the Sun with an uncertainty usually not exceeding 2 minutes of arc.

VI. ACKNOWLEDGMENTS

The authors wish to acknowledge the aid of Mr. R. F. Treharne during the initial stages of this work and the helpful advice and criticism of Dr. J. L. Pawsey throughout the whole of the project.

VII. REFERENCES

ALLEN, C. W. (1947).—*Mon. Not. R. Astr. Soc.* **107** : 386-96.

APPLETON, E. V., and HEY, J. S. (1946*a*).—*Phil. Mag.* **37** : 73-84.

APPLETON, E. V., and HEY, J. S. (1946*b*).—*Nature* **158** : 339.

BAILEY, D. K. (1948).—*Terr. Magn. Atmos. Elect.* **53** : 41-50.

BOLTON, J. G., and STANLEY, G. J. (1948*a*).—*Nature* **161** : 312.

BOLTON, J. G., and STANLEY, G. J. (1948*b*).—*Aust. J. Sci. Res.* A **1** : 58-69.

COVINGTON, A. E. (1948).—*Proc. Inst. Radio Engrs. N.Y.* **36** : 454-7.

LOVELL, A. C. B., and BANWELL, C. J. (1948).—*Nature* **158** : 517-8.

MCCREADY, L. L., PAWSEY, J. L., and PAYNE-SCOTT, RUBY (1947).—*Proc. Roy. Soc.* A **190** : 357-75.

MARTYN, D. F. (1946).—*Nature* **158** : 308.

PAYNE-SCOTT, RUBY, and MCCREADY, L. L. (1948).—*Terr. Magn. Atmos. Elect.* **53** : 429-32.

PAYNE-SCOTT, RUBY (1949).—*Aust. J. Sci. Res.* A **2** : 214-27.

PAYNE-SCOTT, RUBY, and LITTLE, A. G. (1951).—*Aust. J. Sci. Res.* A **4** : 508-25.

PAWSEY, J. L., PAYNE-SCOTT, RUBY, and MCCREADY, L. L. (1946).—*Nature* **157** : 158-9.

PAWSEY, J. L. (1950).—*Proc. Instn. Elect. Engrs.* **97** : 290-310.

RYLE, M., and VONBERG, D. D. (1946).—*Nature* **158** : 339-40.

RYLE, M., and VONBERG, D. D. (1948).—*Proc. Roy. Soc.* A **193** : 98-120.

WALLMAN, H., MACNEE, A. B., and GADSDEN, C. P. (1948).—*Proc. Inst. Radio Engrs. N.Y.* **36** : No. 6, 700-8.

WILD, J. P. (1951).—*Aust. J. Sci. Res.* A **4** : 36-50.

THE POSITION AND MOVEMENT ON THE SOLAR DISK OF SOURCES OF RADIATION AT A FREQUENCY OF 97 Mc/s.

II. NOISE STORMS

By Ruby Payne-Scott* and A. G. Little*

[*Manuscript received July* 11, 1951]

Summary

Using the equipment described in Part I of this series, the apparent position and polarization of the sources of solar radiation at 97 Mc/s. have been measured at some time during about 30 noise storms occurring between May 1949 and July 1950, and the progress of the source during about six of these has been observed in some detail. The best criterion available for the chance of a storm occurring is shown to be the size of the largest visible individual sunspot. The evidence suggests that the storm radiation is associated with this spot only and not with the rest of the group, that the spot magnetic field is the important factor, and that the direction of rotation of the circular polarization of the storm radiation depends on the magnetic polarity of the associated spot, being usually left-handed when the largest spot is a north-seeking pole and right-handed when it is a south pole. Deviations observed between the apparent positions of radio source and visible sunspot on the solar disk are explained in terms of a high origin of the storm radiation in the solar corona, the observed heights varying from 0·3 to 1·0 photospheric radii above the Sun's visible surface.

I. Introduction

Early observers of solar noise at metre wavelengths noted that high-intensity solar radiation lasting for a period of days was usually associated with the passage of a large sunspot group across the solar disk (Hey 1946 ; Pawsey, Payne-Scott, and McCready 1946). Directional observations, based on the interference phenomenon as the Sun rose over the sea, were made in this Laboratory when active spot groups were near the central meridian ; these indicated that the high-intensity radiation observed originated in, or very close to, the area of the sunspot group (McCready, Pawsey, and Payne-Scott 1947). More recently, Wild (1951) has studied the spectrum of such radiation over the frequency range 70–130 Mc/s., showing that it consists of short-lived bursts of narrow bandwidth (of the order of 4 Mc/s.) superposed on an apparently continuous background.

The present paper describes further measurements† on the apparent position and polarization of the radio source during such noise storms, made at 97 Mc/s. with the equipment described in Part I of this series (Little and Payne-Scott 1951). This equipment is a modification of the two-aerial system used by Ryle and Vonberg (1948) ; the main new features are the speed with which measurements can be made, a complete determination of position and polarization taking less than a second, and the more detailed information available on the nature of the polarization. The results obtained on noise storms are discussed in the light of our knowledge of the solar atmosphere and current theories of the origin of solar noise.

---

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

† Some of this work was described to the meeting of U.R.S.I. at Zurich in September 1950.

II. EQUIPMENT AND NOMENCLATURE

The interferometer used (Little and Payne-Scott 1951) enables the polarization of the source to be measured, and its apparent position to be fixed on a line parallel to the hour-circle through the centre of the Sun, over a period of about 2 hours each side of local midday. For high-intensity radiation, a single position and polarization measurement can be made in 1 second. The absolute accuracy in the determination of position is about ±2 minutes of arc for radiation of fairly high intensity. If there is more than one source radiating at the same instant, the position of the weighted mean of the source is obtained. The polarization can be determined as circular, elliptical, linear, or random, or any mixture of these. When radiation of fairly high intensity is received continuously, a measurement of position and polarization is made automatically every minute, with additional measurements if particularly large increases occur.

In reporting the degree and sense of rotation of circular polarization, the convention given by the International Astronomical Union in its instructions for compiling solar-noise data is followed. The polarization is right-handed if the vector in a fixed plane, perpendicular to the ray, rotates clockwise when viewed in the direction of propagation, and the percentage polarization is

$$100[(I_R-I_L)/(I_R+I_L)],$$

where I_R is the intensity received by an aerial disposed to receive right-handed circularly-polarized radiation, and I_L that by one receiving the opposite polarization. The intensities used in this paper are the median values, as defined in the same I.A.U. communication, over a period of about 2 hours on each side of local midday. The " median " is here defined as the value exceeded during half the period considered, and thus is not much affected by large increases of very short duration.

The intensities are taken from an independent record made on 98 Mc/s. This system uses a pair of crossed Yagi aerials, a standard receiver, and an Esterline-Angus recorder. The aerials are switched so as to receive the right-handed and the left-handed circularly-polarized components of the incident radiation in alternate half-minutes, and thus provide a measure of the degree of circular polarization as well as of intensity.

III. EXPERIMENTAL RESULTS

The type of solar radiation with which we are concerned in the present paper is that known as a noise storm (Allen 1947).* In such a storm, the passage of a large sunspot group across the solar disk is accompanied by radio-frequency radiation of higher intensity than normal, lasting for hours and often for days. The polarization of this radiation is normally circular, and the power received on the polarization corresponding to one direction of rotation is usually at least five or six times that corresponding to the opposite direction of rotation. The intensity is by no means constant, and superposed on the slower variations of the background are a series of bursts, the " storm bursts ", each lasting for a few seconds and having the same polarization as the background ; it seems likely that the background itself is composed of numbers of such bursts, only the larger being separately visible (Wild 1951).

---

* This type of radiation has also been referred to as " enhanced radiation " (Payne-Scott 1949 ; Wild 1951).

There is another type of high-intensity solar radiation with which we shall not be concerned in the present paper. This consists of bursts of randomly-polarized radiation, lasting sometimes for a few seconds and sometimes for a quarter of an hour or more. These bursts were first recognized as distinct from " storm bursts " in a paper by Payne-Scott (1949), in which they were described as " unpolarized bursts ", as the equipment then used did not permit of distinction between random and linear polarization. The longer and more intense randomly-polarized bursts are often associated with solar flares and are referred to as " outbursts " (Allen 1947). Those of shorter duration with no accompanying flare or noise storm have sometimes been described as " isolated bursts " (Pawsey 1950). Future papers in this series will deal with the application of the interferometer to these phenomena.

Observations on noise-storm radiation have been made from May 1949 onwards. During each storm some position and polarization measurements have been made, and in six cases of storms of long duration a fairly complete record has been obtained. We shall first discuss the general characteristics of the storm-radiation records, then use the collected interferometer data to try to determine the relationships between noise storms and the associated sunspot groups and finally consider the more detailed relationship between the storm and the spot groups as shown in the six storms observed in detail.

(*a*) *General Characteristics of Storm Radiation*

Figure 1, which is a typical sample of a continuous recording of intensity and circular polarization during a noise storm, shows the characteristics of storm radiation already described. The record is broken up by switching into a series of half-minute intervals in which left-handed and right-handed radiation were received alternately. In this particular case the left-handed component is the stronger. Storm bursts, also showing predominantly left-handed circular polarization, can be seen superposed on the background and often many times greater than it in intensity.

Plate 1 shows a series of typical interferometer records from which position and polarization measurements are computed by the methods described in Part I of this series (Little and Payne-Scott 1951). The appearance of a sine wave in the second frame of each set of three, taken with the two recording aerials mutually at right angles, and the fact that it is shifted in phase by 90°

from the first frame, which is taken with aerials of the same spacing but parallel to one another, indicates that the received radiation is circularly polarized.

Figure 2 shows a histogram of a typical series of position measurements relative to a fixed line on the Sun made with the interferometer during a period of storm radiation such as that shown in Figure 1. Each individual position measurement is the weighted mean of the positions of the source of the continuous radiation and of the source of any storm burst occurring at the moment of measurement. It will be seen that the individual positions varied only slightly about the mean, the scatter being hardly greater than would have been expected from instrumental causes alone, as the probable error is about ±2 minutes of arc. Even when very high storm bursts were present the position did not differ markedly from the mean. The histogram indicates that the apparent position of the source always lies within a strip of width about 6 minutes of arc. It is not

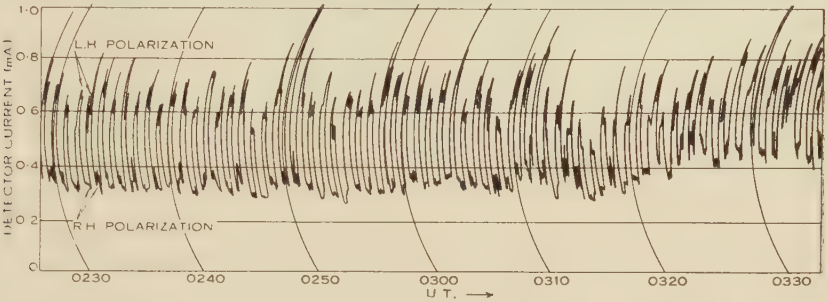


Fig. 1.—Typical recording of storm radiation on 98 Mc/s. Aerials disposed to receive L.H. and R.H. circular polarization in alternate half minutes, L.H. in first half of each minute and R.H. in second half.

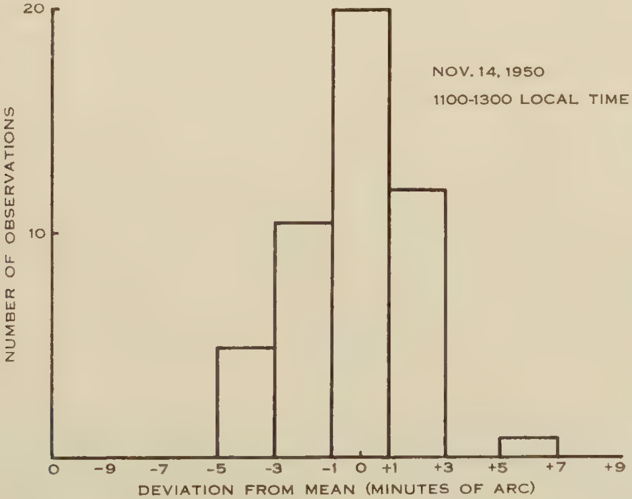


Fig. 2.—Scatter in consecutive observations of source position for source with constant mean position.

possible with the present equipment to determine whether the area of the source of individual bursts is equal to or less than this. They are not much wider, as if they were, the depth of interference pattern from the background with and without large storm bursts superposed would be significantly different, which is not the case, and the area of the continuous background has been determined on a number of occasions by the fixed-lobe interferometer as not greater than 6 or 8 minutes of arc. The scatter in position becomes more marked when the apparent position of the source is close to the solar limb. This is partly due to the fact that the intensity of the radiation is low and errors are thus increased, but there appears to be a real increase in apparent area of the source at such times.

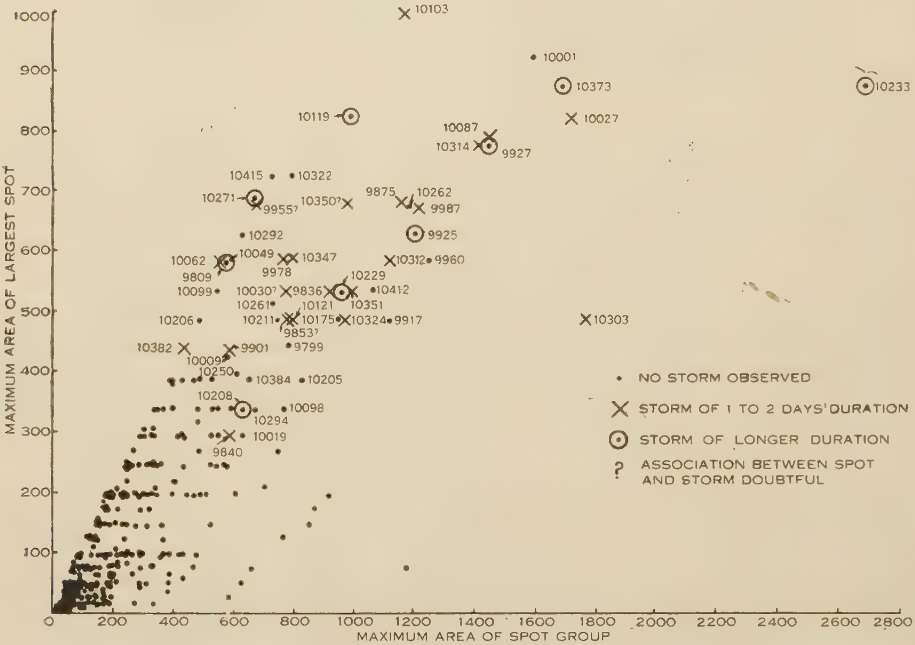


Fig. 3.—Scatter diagram of all spot groups observed at Mt. Wilson during May 1949–July 1950 inclusive, showing noise storms observed.

The observed polarization also usually varies only slightly from minute to minute about a mean value, and that of the background alone does not differ markedly from that of the background with a large burst superposed.

On occasions, marked changes in polarization have been observed, accompanied by changes in position. Such changes were, however, slow, occupying periods of the order of half an hour, and the changes from minute to minute were still small. They are discussed in the last part of this section.

Although we shall not consider in any detail in this paper the behaviour of the randomly-polarized bursts which sometimes occur superposed on storm radiation, we may mention that those occurring during a noise storm appear to arise in the same neighbourhood as the accompanying storm bursts, but may have an apparent position of up to 7 minutes of arc away from the mean apparent position of the storm source, i.e. outside the area indicated by Figure 2. Many randomly-polarized bursts are not accompanied by any detectable noise storm.

On every occasion the observed storm radiation was associated with a large sunspot group. In the next section we shall consider this association in more detail.

(*b*) *Some Relations between Sunspots and Noise Storms*

Between May 1949 and July 1950, some interferometer measurements were made during all noise storms of appreciable size. The lowest level of radiation observed was of the order of 10^{-21} W.m.$^{-2}$ (c/s.)$^{-1}$. As already, mentioned, each storm was associated with a large sunspot group.

Figure 3 is a scatter diagram in which all the spot groups observed at Mt. Wilson during the period May 1949 to July 1950 inclusive have been plotted, using as coordinates the maximum area reached by the group during its passage across the solar disk and the maximum area of the largest spot in the group. When no storm radiation was observed the spot group is denoted by a point, a storm of one or two days' duration is denoted by a cross, and a storm of longer duration by a circle. A question-mark indicates that it is uncertain whether the radiation observed has been related to the correct spot group. The Mt. Wilson spot numbers are given for the larger and the active spot groups.

The scatter diagram demonstrates, in accordance with past observation, that the larger a spot group the greater is the chance that it will be associated with a noise storm. However, it reveals a new fact, that the size of the largest spot in the group is a more certain indication than the size of the whole group of the chance that a noise storm will be associated with the group. Thus a spot group, however large, has very little chance of producing a noise storm of the magnitude considered here unless it contains an individual spot whose area at some time exceeds 400-millionths of the Sun's disk, but as long as the area of an individual spot exceeds this size the area of the spot group is only a secondary consideration. It is known that there is a close relationship between the size of a sunspot and its associated magnetic field, and it is likely that this supplies the reason why size is of such importance.

A further study of large spots which succeed or fail in producing noise storms shows that spots which develop on the solar disk are less likely to produce storms than those which appear round the limb. Thus Table 1, which has been calculated for all spots with an area at some time greater than 400-millionths of the solar disk, shows that a spot which has developed on the disk is only about half as likely to produce a noise storm as one which has appeared round the limb. The early stages in the development of a spot do not seem favourable to the occurrence of storm radiation.

The other information that can be obtained from the collected data on storm radiation is information on the relation between the direction of rotation of circular polarization and the associated spot group. In most cases the direction of rotation remained constant throughout the storm, and was not affected by the passage of the spot group across the Sun's central meridian. However, the direction of rotation of circular polarization appears to be associated with the hemisphere in which the spot occurs. Table 2 shows the relationship between the type of spot, the hemisphere of occurrence, and the type of polarization for all the storms observed.

TABLE 1

RELATION BETWEEN STAGE OF DEVELOPMENT OF LARGE SUNSPOT AND CHANCE OF ASSOCIATION WITH NOISE STORM

Type of Spot	Total Number Associated with			Percentage Associated with		
	Long Storm	Short Storm	No Storm	Long Storm	Short Storm	No Storm
Appeared round limb	7	16	11	21	47	32
Developed on disk ..	1	3	6	10	30	60

TABLE 2

RELATION BETWEEN DIRECTION OF ROTATION OF POLARIZATION OF STORM RADIATION AND HEMISPHERE AND TYPES OF ASSOCIATED SPOT GROUP

Hemisphere and Type*	Number Showing Dominant Polarization as		
	R.H. Circular	L.H. Circular	Variable
N *p*	9	1	—
N *e*	7	—	—
N *f*	—	—	—
S *p*	2	6	2
S *e*	2	1	—
S *f*	1	—	—

* *p*=leader largest ; *f*=follower largest ; *e*=leader and follower about equal.

The first obvious fact is that, although the spots producing storms were almost equally distributed between the two hemispheres, right-handed polarization occurred almost three times as often as left-handed. If the spot group was in the northern hemisphere, the direction of polarization was right-handed in 95 per cent. of the cases in which the leading spot was the largest, or in which the leading and following spots were equal in area. There were no cases in

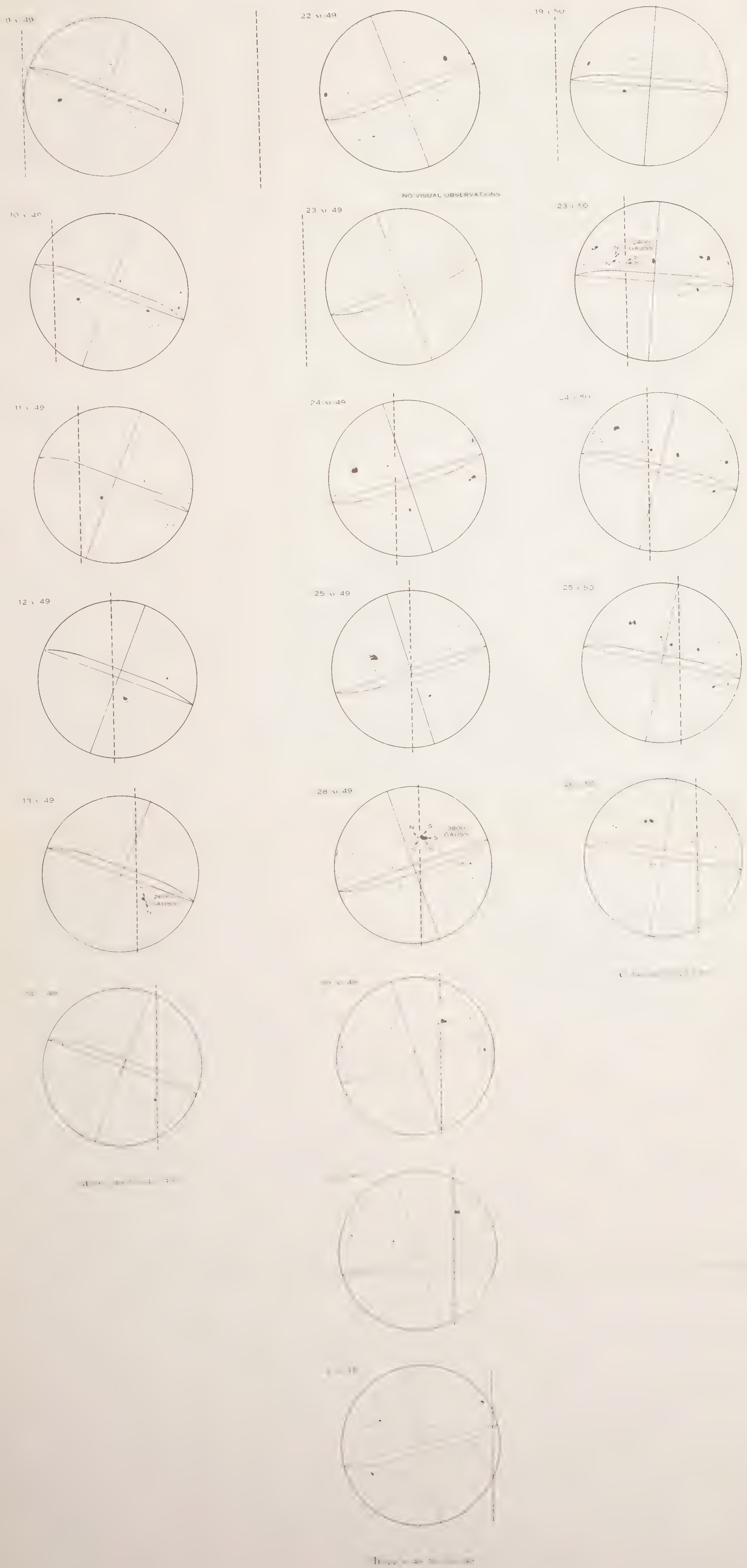
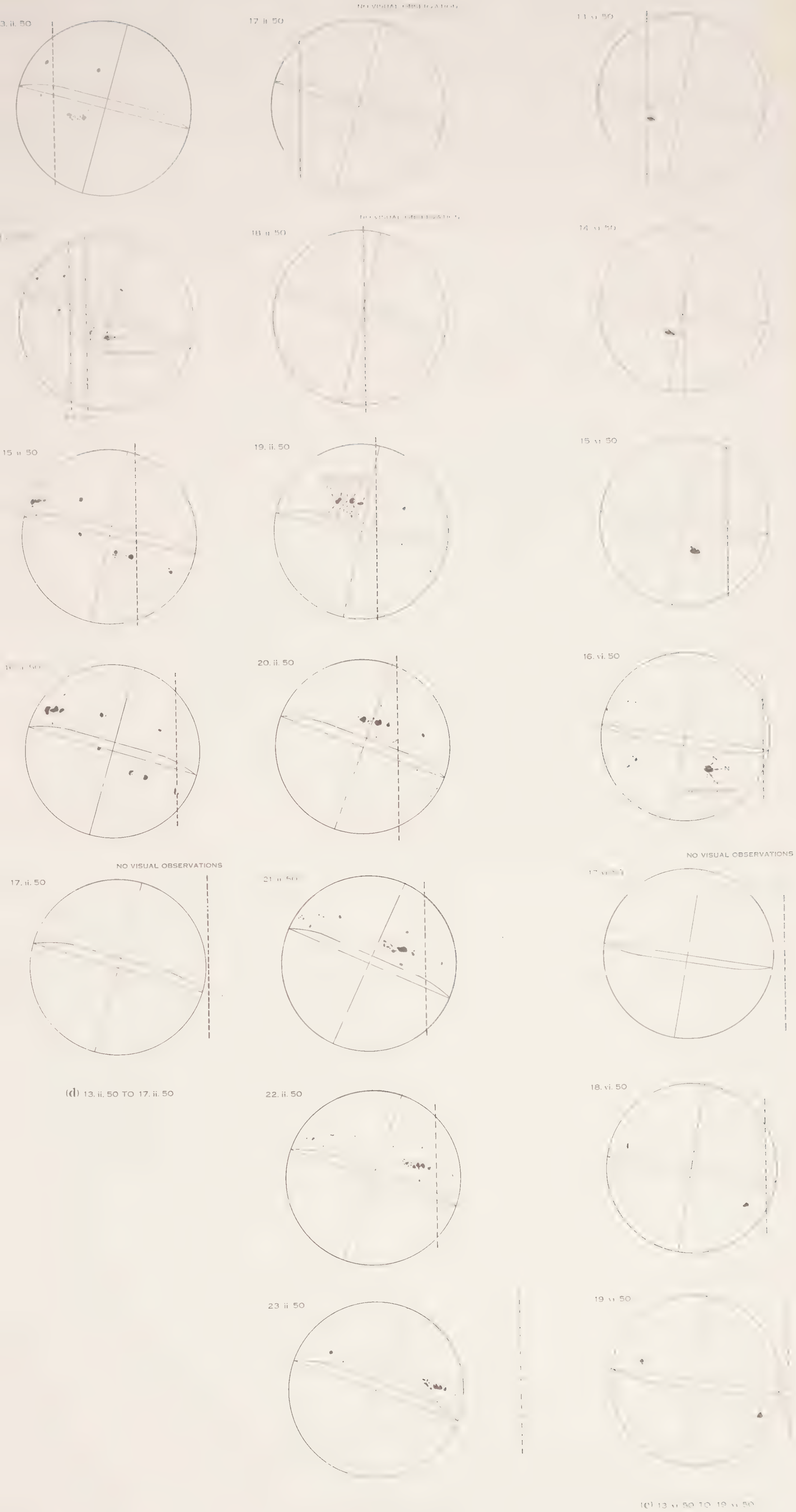

9 x 49
10 x 49
11 x 49
12 x 49
13 x 49
22 xi 49
NO VISUAL OBSERVATIONS
23 xi 49
24 xi 49
25 xi 49
28 xi 49
N
S
S
3800
GAUSS
19 i 50
23 i 50
2400
GAUSS
N
25 i 50
2600
GAUSS

Fig. 4.—Sketches of solar disk during six noise storms showing observed positions of 97 Mc/s source



3. ii. 50
17 ii 50
NO VISUAL OBSERVATIONS
13 vi 50
18 ii 50
NO VISUAL OBSERVATIONS
14 vi 50
15 ii 50
19. ii. 50
15 vi 50
20. ii. 50
16. vi. 50
N
17. ii. 50
NO VISUAL OBSERVATIONS
21 ii 50
NO VISUAL OBSERVATIONS
17 vi 50
22. ii. 50
18. vi. 50
23 ii 50
19 vi 50

(d) 13. ii. 50 TO 17. ii. 50

(d) (CONTINUED) 17 ii 50 TO 23 ii 50

(e) 13 vi 50 TO 19 vi 50

Fig. 4 (continued) Sketches of solar disk during six noise storms, showing observed positions of 97 Mc/s source which the following spot was the largest. If the spot was in the southern hemisphere, then in 70 per cent. of the cases the polarization of the radiation was left-handed rather than right-handed if the leader was larger than or equal to the follower, while in the one case in which the follower dominated, the polarization was right-handed. The significance of these observations is discussed in Section IV.

(c) *The Six Noise Storms Observed in Detail*

A fairly continuous series of observations of position and polarization have been made during six noise storms occurring between May 1949 and June 1950. On each occasion the storm took place while a large spot group was progressing across the solar disk: Figure 4 shows for each storm a sketch* of the spot group for each day when interferometer measurements were made. The mean apparent position of the 97 Mc/s. radiation is shown on each as a line parallel to the observer's meridian. The magnetic polarity* of the elements of the spot group and the value of the maximum central magnetic field* are shown in one of the sketches for each storm. Figure 5 shows the following data in graphical form:

(i) A day-by-day plot of the position of the spot group, in minutes of arc from the central meridian, and of the apparent position of the radio source as observed by the interferometer.

(ii) The median intensity of the radio source for about 2 hours each side of midday on each day, in W.m.$^{-2}$ (c/s.)$^{-1}\times10^{-21}$ at the Earth's surface.

(iii) The mean percentage of circular polarization of the radio source over the same period.

Since the two noise storms observed in February overlapped in time, the plots of position, intensity, and polarization have been combined (Figs. 4 (*d*) and 5 (*d*)). In the second of these storms, huge outbursts of radiation on 97 Mc/s. occurred on February 17 and 22 and were associated with gross changes in position and polarization. Detailed discussion of these is left to Part III of this series, which is devoted to outbursts. During the storm of November 1949 a number of spot groups were present on the solar disk, and the radiation on November 24 and 25 appears to have a different origin from that during the remainder of the storm. The direction of polarization on those days is also opposite to that during the rest of the period.

The experimental results may be summarized as follows:

(i) *Relation to Spot Group.*—The radiation in each case is associated with the largest spot group visible at the time, and travels across the Sun with the group. The apparent position of the radio source approximately coincides with that of the associated spot group near central meridian passage, but the source appears to move faster than the spot group, so that its position becomes displaced towards the solar limb relative to that of the spot group. At the beginning of each noise storm (which always occurs after the spot group first becomes visible) and at the end (which always occurs before the spot group disappears) the displacement becomes much greater, so that the apparent source of the radiation is often off the solar limb, sometimes by as much as a solar radius (November 22, 1949, and February 23, 1950). In the case of the storm of June 1950 (Figs. 4 (*e*) and 5 (*e*)) the radiation, after travelling out to the solar limb, remained in roughly the same position for the last three days. In only one of the storms shown here, i.e. that of February 17–23, were outbursts accompanying flares observed, and in this storm the progress of the source was more irregular than in the others shown. The rate of movement across the solar disk appears to increase with the intensity of the storm, being greater in storms

---

* The sketches and polarity data were kindly made available by the Commonwealth Observatory, Canberra. The value of magnetic field is taken from the Mt. Wilson data published in the *Publications of the Astronomical Society of the Pacific*.

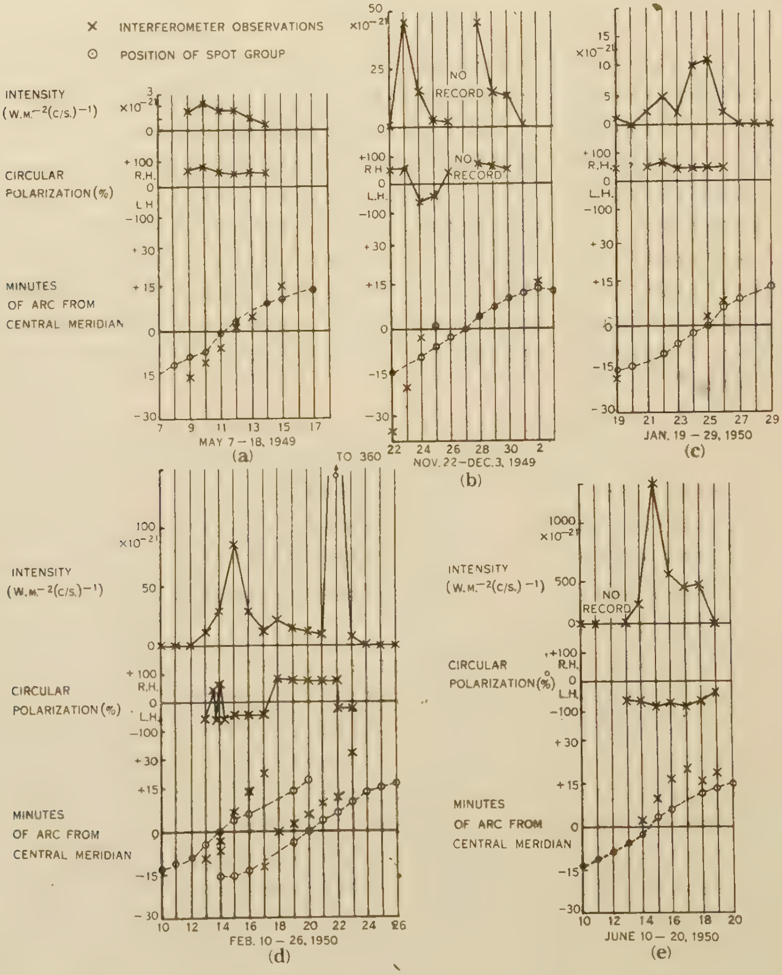


Fig. 5.—Observations on six noise storms, showing daily values of intensity, percentage of circular polarization, and apparent position of radio source and of the associated spot group.

of February 13–17 and June 13–19 (Figs. 4, 5 (*d*), and 5 (*e*)) than in the other four storms.

(ii) *Polarization.*—The radiation during a noise storm usually consists of an admixture of right- and left-handed circular polarization (or of random and circular polarization, the two methods of expressing the admixture being equivalent if no consistent phase difference is assumed between right- and left-handed components), with one sense of circular polarization predominant. Elliptical and linear polarization have only been observed at times when the polarization is changing.

Often the polarization remains constant in sense throughout the noise storm, and there is no evidence for a change in polarization at C.M.P. of the spot groups, as suggested by Martyn (1946). However, there is, as we have already seen, a relationship between the hemisphere in which the spot group occurs and the polarity of the radiation. In the storms of Figures 5 (*b*) and 5 (*c*) and the second storm of 5 (*d*) the polarization was predominantly right-handed and the associated spot groups were in the northern hemisphere, while in the first storm of 5 (*d*) and the storm of 5 (*e*) the storm was predominantly left-handed and the spot groups were in the southern hemisphere. The exception was the storm of Figure 5 (*a*), in which a spot group in the southern hemisphere produced only right-handed polarization. In this spot group the following spot was much the larger, while in the others either the leading spot was the largest or the leader and follower were much the same size (Fig. 4). In this storm also the apparent position of the source tends to lag behind the spot group, while in the others it is, on the average, on or ahead of the spot group near C.M.P.; the displacements in all cases are not greater than the probable error, and may not be significant.

Of the four changes in polarization occurring during the noise storms shown, two were associated with outbursts, and will be discussed in Part III of this series. The other two changes occurred during the storm in November-December 1949, and on February 14, 1950. In the first, the changes in polarization from right to left and back to right occurred out of observing hours, and appear to have been due to a change in the emitting source from one spot group to another. On February 14, interferometer measurements were made from 0028 to 0337 U.T. Earlier, the polarization had been left-handed. When measurements began, the polarization was markedly elliptical, approaching linear, and was right-handed. The ellipse gradually approached closer to a circle, the direction of polarization still being right-handed. Between 0230 and 0300 the polarization changed again. Unfortunately no polarization measurements were made with the interferometer during this period, but the 98 Mc/s. record shows the polarization as either random or linear. From 0300 onwards, the polarization was left-handed and almost completely circular. The mean positions of the radiation of different polarizations, in minutes of arc west of the hour-circle through the centre of the solar disk, were 6 minutes for the period in which the polarization was linear, 7 minutes for the period during which it was right-handed and approaching circular, and 4 minutes for the later period in which it was left-handed circular. The difference between these positions is again of the same order as the experimental error.

(iii) *Time of Maximum Intensity.*—Allen (1947) deduced from his 200 Mc/s. observations that there is a sharp peak in the intensity of storm radiation within a day of central meridian passage of the associated spot group. His curve, based on 24 noise storms, is reproduced in Figure 6, together with a similar curve derived from Figure 5, in which the ratio of intensity to that at C.M.P. was used rather than absolute intensity to prevent the curve being dominated by the largest storm; the storm of November-December 1949 was omitted, as there were no measurements near C.M.P. The third curve in Figure 6 is a cosine curve giving the expected change in intensity if the decrease in intensity away from C.M.P. is due solely to the decrease in projected area of the spot group.

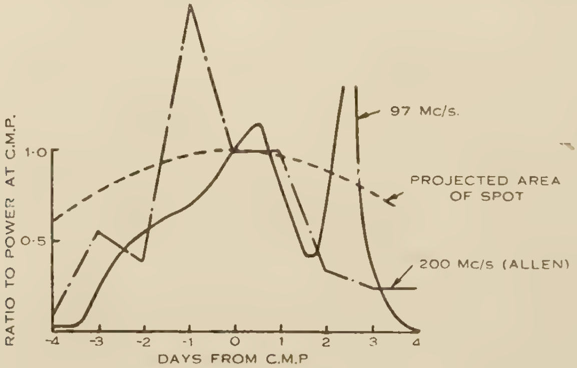


Fig. 6.—Intensity of storm radiation related to distance of associated spot group from central meridian passage.

It would seem that, although there is a general tendency for the received power to be greatest near C.M.P. this is often offset at 97 Mc/s., by local variations in intensity, one of which gives rise to the high peak at about $+2\frac{1}{2}$ days in the curve shown. The displacement of the peak west of C.M.P. is also mainly due to one storm. It is clear, however, that the power received more than 2 days away from C.M.P. is usually much lower than would be expected if the decrease were due solely to decrease in the projected area of the spot group, so that the received radiation appears to have some measure of directivity.

Another important factor in determining the time of maximum intensity appears to be the size of the largest sunspot in the associated group. Table 3 shows for each storm the date of the maximum intensity, the date on which the area and the projected area* of the largest visible spot were greatest, and the date of C.M.P. of the largest spot. Since the intensity measurements refer to a period between 0000 and 0400 hours U.T. on the date given, while the spot areas are measured at about 1500 hours U.T., a discrepancy of 1 day between dates of maximum intensity and maximum spot area is hardly significant. It will be

---

[*] Spot areas throughout this paper are taken from the circulars issued by the United States Naval Observatory.

seen that the time of maximum intensity is on the whole closer to the time when the largest spot had its maximum area than to the time of C.M.P. The low intensities observed during central meridian passage of the second large group of February 1950 seem to be associated with the fact that at this stage the group contained no single large spot, while the high level on February 22 occurred at about the time when a number of small spots combined to form one large one. The one case in which a higher intensity occurred at the time of C.M.P., rather than at the time of greatest spot area, was that of January 1950. This storm was also exceptional in other ways, as the areas of both the spot group and the largest spot, and also the value of the central magnetic field, were smaller than for the other storms. This spot reappeared as the very large spot of late February, so that its capacity to produce storm radiation may have been related to its capacity for development.

Table 3

DATE OF MAXIMUM STORM INTENSITY RELATED TO DATE OF (*a*) MAXIMUM SPOT AREA, (*b*) MAXIMUM PROJECTED AREA, AND (*c*) C.M.P. OF LARGEST SPOT, FOR SEVERE STORMS

Date of Storm	May 1949	Nov.-Dec. 1949	Jan. 1950	Feb. 1950 (1)	Feb. 1950 (2)	June 1950
Max. intensity ..	May 10	Nov. 23 and 28	Jan. 25	Feb. 15	Feb. 22	June 15
Max. spot area ..	May 10	Nov. 24	Jan. 21	Feb. 16 and 17	Feb. 16 and 23	June 19
Max. projected area	May 10	Nov. 28	Jan. 21	Feb. 16	Feb. 16 and 22	June 15 and 19
C.M.P. of largest spot	May 11·5	Nov. 27	Jan. 24·5	Feb. 14	Feb. 20	June 14

IV. Interpretation of Results

(*a*) *The Connection between Spot Size and the Occurrence of Noise Storms*

The evidence given here indicates that storm radiation is associated with only one spot in a group—the largest. Since it is known that there is an empirical relationship between spot size and magnetic field, and since the circular polarization observed in noise storms suggests that the spot field must reach at least to the level of origin of the radiation, it seems likely that the true criterion for the production of storm radiation is the presence of a magnetic field of adequate strength at an appropriate level in the corona. Unfortunately no information is available on the value of spot fields in the corona and it does not seem likely that such measurements will be made by optical means. The average value of the central magnetic field at the photosphere of the spot group for the groups associated with storms was 3300 gauss.

The apparent higher origin in the corona of the two storms of highest intensity from the six observed in detail suggests that the spot field on these occasions may have reached greater heights than in the other storms. The increase in intensity may possibly be due to reduced absorption for radiation produced higher in the solar atmosphere.

The observations indicate that spots which develop during their passage across the solar disk, even though they may rapidly develop to a large size, are less likely to produce noise storms than those which appear round the eastern limb of the Sun. If the necessary condition for the production of storm radiation is the existence of a sufficient magnetic field at appropriate levels in the corona, then this observation suggests that the magnetic field in such developing spots takes some days to be established in the corona.

(*b*) *Polarization*

The apparent association of storm radiation with only one spot in a group makes it easier to explain the usual predominance of one direction of rotation of polarization. In any one sunspot cycle leading spots in bipolar groups have one polarity in one hemisphere and the opposite polarity in the other, all the polarities reversing in the next cycle. In the present cycle the leading spots in the northern hemisphere are south-seeking magnetic poles, and in the southern hemisphere are north-seeking magnetic poles, as may be seen from Figure 3. Thus we might expect that storm radiation would have polarization of one direction of rotation of polarization if it were associated with a leading spot in the northern hemisphere, or a following spot in the southern hemisphere, and the opposite direction if associated with a leading spot in the southern hemisphere, or a following spot in the northern hemisphere. Table 2 shows that such a rule was followed by 75 per cent. of the cases in which the leading spot was largest, and also by the only case in which a following spot was largest, the storm radiation associated with a south magnetic pole showing right-handed circular polarization and that associated with a north pole showing left-handed circular polarization. In addition, there appears to be a tendency over the period of the present observations towards the production of right-handed circular polarization.

The same association between the direction of rotation of polarization and the polarity of the associated spot is also suggested by the observations of the relation between the apparent position of the source and of the large spot in relation to the rest of the group, although the position differences concerned are of the same order as the experimental error and may be in doubt. Thus Figure 4 shows that in the storm for which radiation appears to be associated with the following spot (May 1949) the position of the radiating source lags behind the centre of the spot group even near C.M.P., while in the other five storms it tends rather to be ahead of the centre of the group. The observed position changes on February 14 (Section III (*c*)) suggest that a change from right-handed to left-handed circular polarization was accompanied by a change in emitting source from above a south magnetic pole to above a north magnetic pole, while the opposite polarization change was accompanied by the opposite movement; the linear polarization appears to have been produced by a source midway between the two poles of the spot.

If the observed direction of rotation of polarization is to be explained on magneto-ionic theory as the limiting polarization received as the radiation emerges through regions of low density in the solar atmosphere, then for propagation with a component in the direction of the field the direction of rotation should be left-handed for the ordinary ray and right-handed for the extraordinary ray, the directions being reversed when the direction of propagation is reversed. Thus the observed association between right-handed circular polarization and a south polar spot, and conversely left-handed and a north polar spot, would suggest that the radiation reaching the Earth is predominantly the ordinary ray, if the magnetic field concerned is that due to the sunspot and has a conventional distribution. Linear polarization would be expected then only when the field was strictly normal to the direction of propagation, and such a condition would seem unlikely. However, in addition to the case described here, two further instances of linear polarization which occurred during outbursts will be described in the next paper of this series. In both cases the direction of the electric vector was approximately along the axis of the spot group, which is the direction of polarization to be expected for the ordinary ray if the magnetic field involved is parallel to the axis of the spot group.

(c) *The Divergence between Apparent Positions of the Radio Source and of the Associated Spot Group*

It has been shown (Ginsburg 1946 ; Martyn 1948) that radiation at radio frequencies cannot emerge from the level of the Sun's photosphere, but must be emitted from higher levels in the Sun's atmosphere. Hence even if the storm radiation originates radially above the associated spot group, we should expect a discrepancy between the observed positions (projected onto the solar disk) of the radio source and the associated spot group, except when both are on the central meridian. We shall consider whether the observed differences in position can be completely explained by such a difference in height.

Figure 7, taken from a paper by Jaeger and Westfold (1950), shows the computed paths of a bundle of parallel rays of 100 Mc/s. radiation entering the solar atmosphere from the outside, in the absence of a magnetic field, using the conventional distribution of electron density in the corona. Each ray is bent and is eventually reflected, emerging again along another path, the turning point occurring when the refractive index is zero for normal incidence and earlier the more oblique the angles of incidence. The locus of turning points is a curve whose distance from the Sun's visible surface is about $0 \cdot 15$ photospheric radii for normal incidence, increasing as the ray becomes more oblique. Conversely radiation emerging from the atmosphere will follow these paths in reverse, and radiation arising inside the locus of turning points cannot escape towards the Earth. In the presence of a magnetic field (e.g. in the neighbourhood of a sunspot) a ray entering the solar atmosphere is split into two rays, the ordinary and extraordinary, for which the medium has different refractive indices. The locus of turning points for ordinary rays has the same central point as, and will not differ greatly from, that of Figure 7, while for the extraordinary ray we have two similar loci, one above and one below that for the ordinary ray. Smerd (1950), using the conventional values for coronal electron densities, has calculated that for a large sunspot with a field of 3600 gauss the outer level of zero refractive index for the extraordinary rays is at a height of 0·3 photospheric radii above the Sun's visible disk.

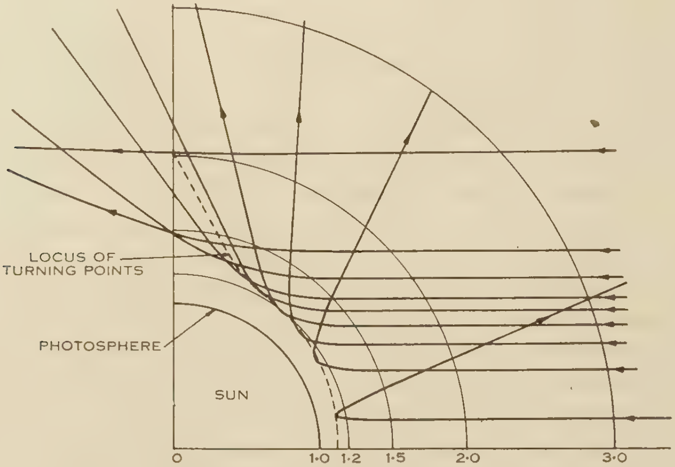


Fig. 7.—Ray paths at 100 Mc/s. in solar atmosphere in absence of magnetic field (from Jaeger and Westfold 1950).

It is usually assumed that the extraordinary ray cannot escape through this outer level, though Saha (1947) and Bailey (1950) have suggested that under some circumstances such an escape is possible. We shall see that the observations described here indicate that at this frequency storm radiation is not emitted from below this outer level.

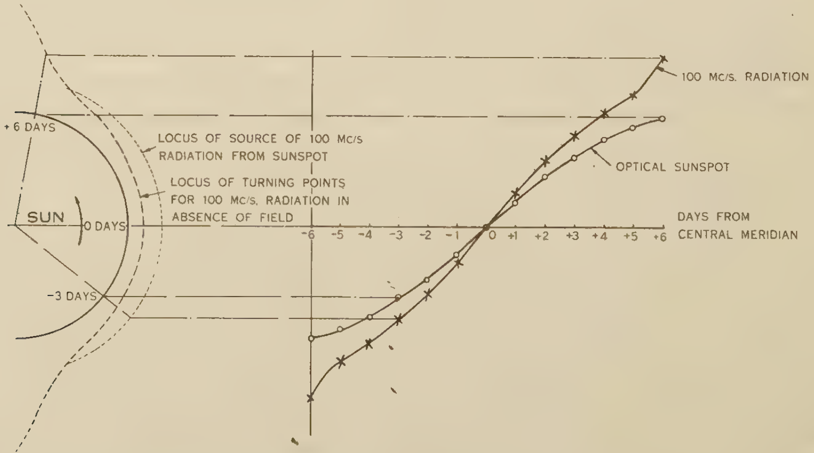


Fig. 8.—Relation between apparent positions of 100 Mc/s. source and of associated spot group, assuming 100 Mc/s. source in corona.

Let us then assume that the source of noise-storm radiation is at a height of, say, 0·3 photospheric radii above the Sun's visible surface, and is always radially above the associated spot group. Figure 8 shows this postulated path, as well as the locus of limit of escape of Figure 7. When the spot group is near either limb of the Sun, the assumed path of the storm radiation will fall inside the locus of Figure 7, and radiation can no longer emerge from this region. Since the source of radiation is presumably extended, some radiation will still be received, probably of much lower intensity from levels in the solar atmosphere at or beyond the locus of turning points. The graph on the right-hand side of Figure 8 shows the relation between the apparent positions on the solar disk of the spot group and of the radio source on these assumptions.

The behaviour of the radio source seen in this graph exhibits the characteristic features observed in Figures 4 and 5. Over the main part of the storm, the ratio between the apparent distances of radio source and spot from the central meridian remains fairly constant at a value greater than unity, while at the beginning and end a considerable drop in intensity is often accompanied by large deviations towards the solar limb. Also, storms always begin some time after the associated spot group first appears round the eastern limb of the Sun, and disappear before the spot disappears round the western limb, because most of the emitting source disappears into levels from which radiation cannot escape.

Quantitatively, the observations require the level of origin of the radiation to be about 0·3 or 0·4 radii above the Sun's visible surface for four of the storms shown, and from 0·8 to 1·0 radii above for the two storms of highest intensity (Figs. 5 (*d*), first storm, and 5 (*e*)). The deviations from a smooth curve often observed in the apparent position of the storm radiation may be due to changes in the height of the emitting source. The actual duration of a storm is usually considerably less than would be anticipated from Figure 8, as can be seen by comparing the period during which the assumed locus of the source of 100 Mc/s. radiation lies outside the locus of the limit of escape with the actual duration of storms as shown in Figures 5 and 6. This discrepancy between theory and experiment is removed if we assume that the locus of the limit of escape is higher in the corona than shown. There is evidence from measurements at metre wavelengths of the thermal radiation from the Sun that the conventional values for coronal electron densities should be increased, at least near the maximum of the sunspot cycle, by a factor of the order of 2 (see, for example, Waldmeier 1951). The resultant increase in height in the locus of the limit of escape in Figure 8 will result in a considerable decrease in storm duration even when the locus of 100 Mc/s. radiation is also increased to that deduced from the observations.

We have ignored the possibility of refraction as a primary reason for the position changes. Inspection of Figure 7 shows that in a uniformly stratified atmosphere refraction is very small, and results in the source being displaced towards the centre of the Sun. We could obtain bending in the opposite direction by assuming suitably shaped regions of very high density, but it seems unnecessary to invoke such special regions when, as we have seen, the observations can be satisfactorily explained by assuming a source high in the corona at a level of between 0·3 and 1·0 photospheric radii above the photosphere. These levels are of the same order as those deduced for " unpolarized bursts " (Payne-Scott 1949). They are above those calculated for the level of reflection for the ordinary ray, and at or above that for the extraordinary ray calculated from a conventional distribution of electron density and spot magnetic field (Smerd 1950). We have seen that it is likely that these calculated levels should be increased, but it still seems probable that, at least for the more intense noise storms, the source of radiation lies beyond the level of reflection for the ordinary ray.

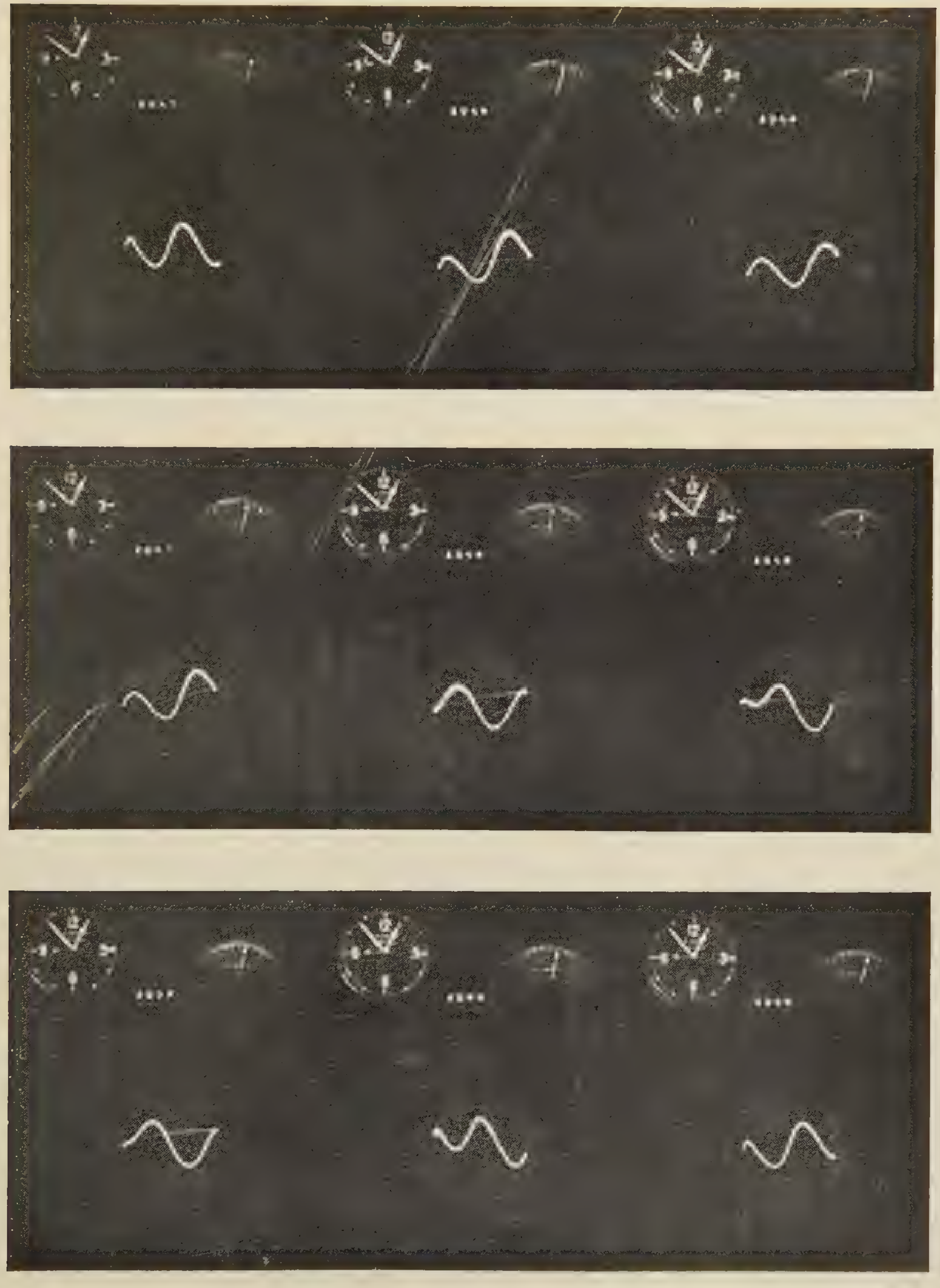
V. CONCLUSIONS

The observations quoted here point to magnetic fields as the important factor in the production of solar noise storms at a frequency of 97 Mc/s. The apparent association of spot radiation with only one spot in a spot group—the largest—can, owing to the connection between spot size and spot field, be interpreted as evidence that a certain value of magnetic field is required for the production of storm radiation of the intensities observed. The observed directions of rotation of polarization suggest that a south-seeking magnetic pole is associated with right-handed polarization and a north-seeking pole with left-handed. The picture appears to be complicated by a general tendency towards right-handed polarization during the observing period. Since the polarities of spots will reverse in the next few years as spots of the new sunspot cycle appear, there will soon be a good opportunity to test the connection between the direction of rotation of polarization and the polarity of the associated sunspot.

The deviations in position between the apparent position of the solar source and that of the associated sunspot can be explained by assuming a source of origin high in the corona, at a height ranging from 0·3 to 1·0 photospheric radii above the Sun's visible surface, and hence usually above the level of zero refractive index for either the ordinary or extraordinary ray of magneto-ionic theory calculated by assuming a unipole field above a spot and a conventional distribution of electron density. The evidence on the height and duration of storms suggests that the accepted values for coronal densities should be increased, and in fact the observation of storm radiation may provide data on electron densities and magnetic fields in the corona at levels where optical methods fail. The observation that spots developing rapidly on the solar disk are less likely to produce storm radiation than those which appear round the eastern limb may throw some light on the rate at which magnetic fields develop in the corona.

Unfortunately most of the theories put forward to explain the origin of storm radiation do not lead to any definite predictions of region of origin, relation to sunspots, or polarization, and hence are not capable of verification or refutation by the observations recorded here. The high observed level of origin appears to rule out (for these frequencies) those theories which give sources low in the solar atmosphere, such as that of Saha (1947) and that portion of the theory of growing waves (Bailey 1950) which gives a low level of origin. The requirement seems to be a mechanism producing high-intensity radiation in a region of presumably very low density with a fairly high magnetic field. The reason why such a region should emit more efficiently than lower regions in the solar atmosphere may be related to the much lower absorption experienced by radiation from

MOVEMENT OF SOURCES OF RADIATION ACROSS THE SUN'S DISK. II



Typical interferometer records made on film, with 25 c/s. time base.

such regions. This would account for the greater height of origin observed for the more intense storms.

The absorption in such a region would be lower for the ordinary ray than the extraordinary, and this might account for the suggested predominance of the ordinary ray in the radiation received. The exciting mechanism must then, of course, be non-thermal, otherwise the low absorption will be balanced by low emission; Ryle (1948) has suggested that storm radiation is due to local temperature increases, but the heights of origin observed seem much greater than his theory would indicate and the polarization appears to be in the wrong direction.

VI. ACKNOWLEDGMENT

The authors wish to thank Dr. J. L. Pawsey for helpful advice and criticism throughout the whole of the project.

VII. REFERENCES

ALLEN, C. W. (1947).—*Mon. Not. R. Astr. Soc.* **107**: 386-96.

BAILEY, V. A. (1950).—*Phys. Rev.* **78**: 428-43.

GINSBURG, V. L. (1946).—*C.R. Acad. Sci. U.R.S.S.* **52**: 487-90.

HEY, J. S. (1946).—*Nature* **157**: 47-8.

JAEGER, J. C., and WESTFOLD, K. C. (1950).—*Aust. J. Sci. Res.* A **3**: 376-86.

LITTLE, A. G., and PAYNE-SCOTT, RUBY (1951).—*Aust. J. Sci. Res.* A **4**: 489-507.

MCCREADY, L. L., PAWSEY, J. L., and PAYNE-SCOTT, RUBY (1947).—*Proc. Roy. Soc.* A **190**: 357-75.

MARTYN, D. F. (1946).—*Nature* **158**: 308.

MARTYN, D. F. (1948).—*Proc. Roy. Soc.* A **193**: 44-59.

PAWSEY, J. L. (1950).—*Proc. Instn. Elect. Engrs.* **97**: 290-310.

PAWSEY, J. L., PAYNE-SCOTT, RUBY, and MCCREADY, L. L. (1946).—*Nature* **157**: 158-60.

PAYNE-SCOTT, RUBY (1949).—*Aust. J. Sci. Res.* A **2**: 214-27.

RYLE, M. (1948).—*Proc. Roy. Soc.* A **195**: 82-97.

RYLE, M., and VONBERG, D. D. (1948).—*Proc. Roy. Soc.* A **193**: 98-120.

SAHA, M. N. (1947).—*Indian J. Phys.* **21**: 199-221.

SMERD, S. F. (1950).—*Proc. Instn. Elect. Engrs.* **97**: 447.

WALDMEIER, M. (1951).—*Naturw.* **38**: 1-4.

WILD, J. P. (1951).—*Aust. J. Sci. Res.* A **4**: 36-50.

A HIGH SPEED CHRONOGRAPH AND ITS APPLICATION TO THE TIME OF FLIGHT OF COSMIC RAYS

By V. C. Officer*

[*Manuscript received July 2, 1951*]

Summary

A high speed chronograph using a 20 turn spiral time base of 1 μsec. duration per turn is described, together with its video amplifiers, delay lines, radial deflection circuits, and recording equipment. The time of flight of cosmic ray mesons over a range of 5·45 m. has been detected using ordinary Geiger counters, and found to agree with the velocity of light within the accuracy of the experiment. A time lag of 2×10^{-8} sec. between the firing of a tray of counters shielded with 10 cm. of lead and the firing of a similar unshielded tray by extensive air showers has been accounted for, and the results shown to be consistent with the theoretical expectation that most of the shower particles arrive at the trays within a time $<10^{-9}$ sec.

I. Introduction

A cathode ray tube chronograph employing a spiral time base has the advantage of not requiring periodic calibration if the spiral deflection voltage is derived from a crystal oscillator. An equally important advantage is its ability to measure both short and long time intervals with high accuracy. In the instrument to be described, one turn of the 20 turn spiral is completed in 1 μsec., and is 20 cm. long, the whole trace being in the neighbourhood of 300 cm., the exact length depending on the pitch chosen. A method of producing radial deflection has been developed that enables a time interval to be marked by radial blips lasting about 2×10^{-8} sec. Measuring between the peaks of these blips gave a standard deviation of $5\cdot0\times10^{-9}$ sec. for a group of 47 observations of a constant 0·410 μsec. time interval occurring at random in any angular position on the first turn.

The instrument was originally used for μ-meson lifetime measurements, but it has now been applied to the study of other short time intervals. The chief source of error in any of these observations arises from the fluctuations in the reaction times of the Geiger counters used to detect the events which define the time intervals. The difference between two randomly chosen reaction times always appears as an error in any single observation of a time interval. Figure 1 shows a typical distribution of these errors when wide band amplifiers are used to transmit the pulses from Geiger counters of conventional proportions.

According to the statistical theory of large samples, the standard error in the sample mean is the sample standard deviation divided by the square root of the number in the sample. Then with a sample of 100 observations of a time interval these counters would yield a result having a standard error of

---

* Physics Department, University of Melbourne.

$\pm 0 \cdot 6 \times 10^{-8}$ sec.* The possibility of detecting the time of flight of cosmic ray particles over short distances is immediately apparent, for at the velocity of light 2×10^{-8} sec. is required to traverse 6 m.

II. The Chronograph

(*a*) *General Outline*

Figure 2 shows a block diagram of the circuit. A circular time base is derived from a 1 Mc/s. oscillator in the usual way, by splitting its output voltage into a four phase supply, each phase being separated by 90° from the next, and the four signals after amplification to an amplitude of 500 V., applied to the deflection plates of the 908PW cathode ray tube. The cathode ray tube is a 9 cm. diameter General Electric type designed for a 10 kV. accelerating voltage and a maximum single stroke photographic writing speed on the tube face of 3000 km/s. In the quiescent condition the cathode ray beam is switched off, and the film in the shutterless camera continuously exposed to the screen by means of an f : 1 bloomed lens. A Wratten 38A blue filter eliminates most of the cathode heater light.

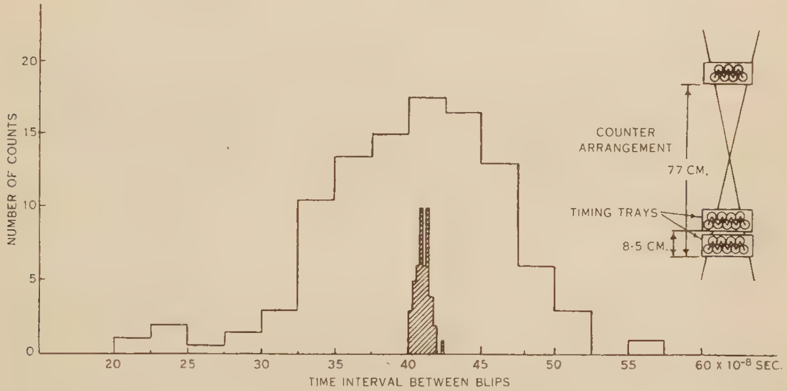


Fig. 1.—Distribution of 104 relative time lags for overlapped 2 cm. diameter counters with central wire 0·01 cm. in diameter, filled to 13 cm. of Hg with a 10 per cent. ethylene : 90 per cent. argon mixture and operated at 1080 V. The standard deviation is $6 \cdot 2 \pm 0 \cdot 4 \times 10^{-8}$ sec. The distribution of 47 observations of the time origin is shown shaded.

When the coincidence unit detects the fact that an event to be timed has occurred, its output pulse switches on the cathode ray beam, and triggers the modulating pulse which causes the circular time base to diminish in amplitude, thus generating a spiral. At the end of this modulating pulse the camera motor is energized and the film wound on.

The timing information is derived from the discharge of Geiger counters connected to the two timing channels. In each channel a cathode follower couples the pulse into a long cable leading to a video amplifier. The output pulses from the two channels are then clipped to 0·1 μsec. duration, mixed, and passed through the same delay line. Possible variations in the delay of this line are therefore of no importance. The delay of 0·35 μsec. introduced, gives the coincidence circuit time to develop an output and switch on the cathode ray beam. After a final amplification the timing pulses are fed to the radial deflection circuit.

---

* Small diameter counters are now being developed and it is expected that their reaction times will be very much less than those of the above counters.

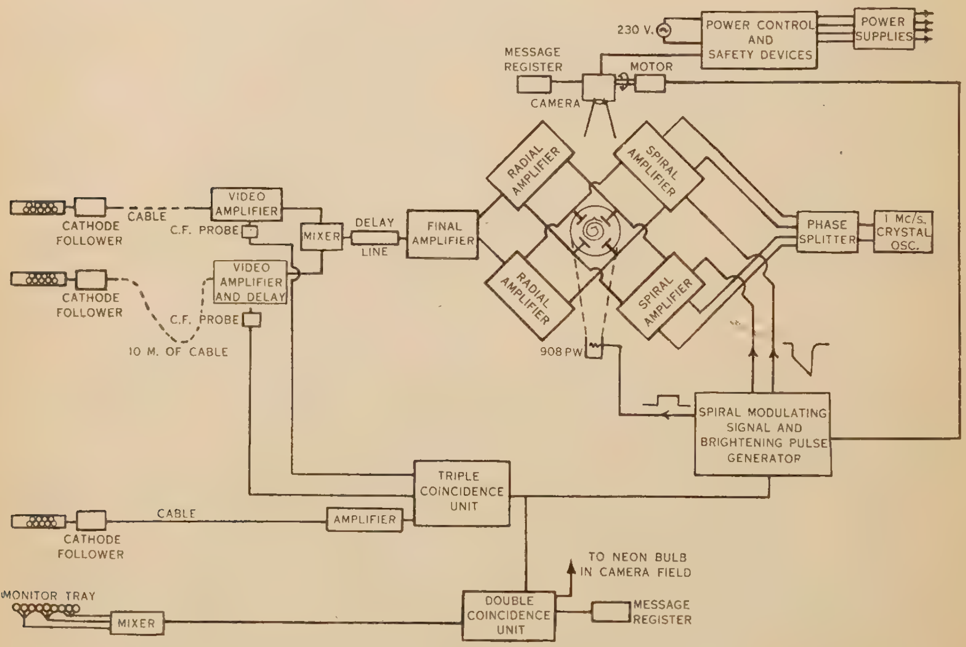

230 V.
POWER CONTROL AND SAFETY DEVICES
POWER SUPPLIES
MESSAGE REGISTER
MOTOR
CAMERA
RADIAL AMPLIFIER
SPIRAL AMPLIFIER
VIDEO AMPLIFIER
CABLE
CATHODE FOLLOWER
C.F. PROBE
MIXER
DELAY LINE
FINAL AMPLIFIER
PHASE SPLITTER
1 Mc/s. CRYSTAL OSC.
VIDEO AMPLIFIER AND DELAY
CATHODE FOLLOWER
C.F. PROBE
10 M. OF CABLE
RADIAL AMPLIFIER
SPIRAL AMPLIFIER
908 PW
SPIRAL MODULATING SIGNAL AND BRIGHTENING PULSE GENERATOR
TRIPLE COINCIDENCE UNIT
CABLE
AMPLIFIER
CATHODE FOLLOWER
TO NEON BULB IN CAMERA FIELD
MONITOR TRAY
MIXER
DOUBLE COINCIDENCE UNIT
MESSAGE REGISTER

Fig. 2.—Block diagram of chronograph.

(*b*) *Radial Deflection*

The method of producing radial deflection developed, allows higher speed to be used than does the usual electronic method, in which the circular time base radius is varied according to the instantaneous signal voltage as well as the slower spiral producing pulse. This latter system demands that the deflection amplifier handle the very large amplitude but narrow bandwidth spiral time base signal, and the small amplitude but extremely wide bandwidth radial deflection signal at the same time. This is extremely difficult to achieve. In the present system the two functions are separated, a radial deflection amplifier being modulated by the spiral signals, and the spiral and radial signals mixed at the cathode ray tube deflection plates.

A pulse carrying timing information to be displayed as a radial deflection, emerges from the final video amplifier (Fig. 3) and switches off the current in four 807 valves, one connected to each deflection plate of the cathode ray tube. In their plate circuits these valves have 2·5 μH. inductances approximately critically damped. Voltage pulses lasting about 0·02 μsec. are developed across these inductances when the current is switched off. The current that is switched off, is in each tube made proportional to the 1 Mc/s. spiral generating voltage appearing on the deflection plate connected to that particular tube.

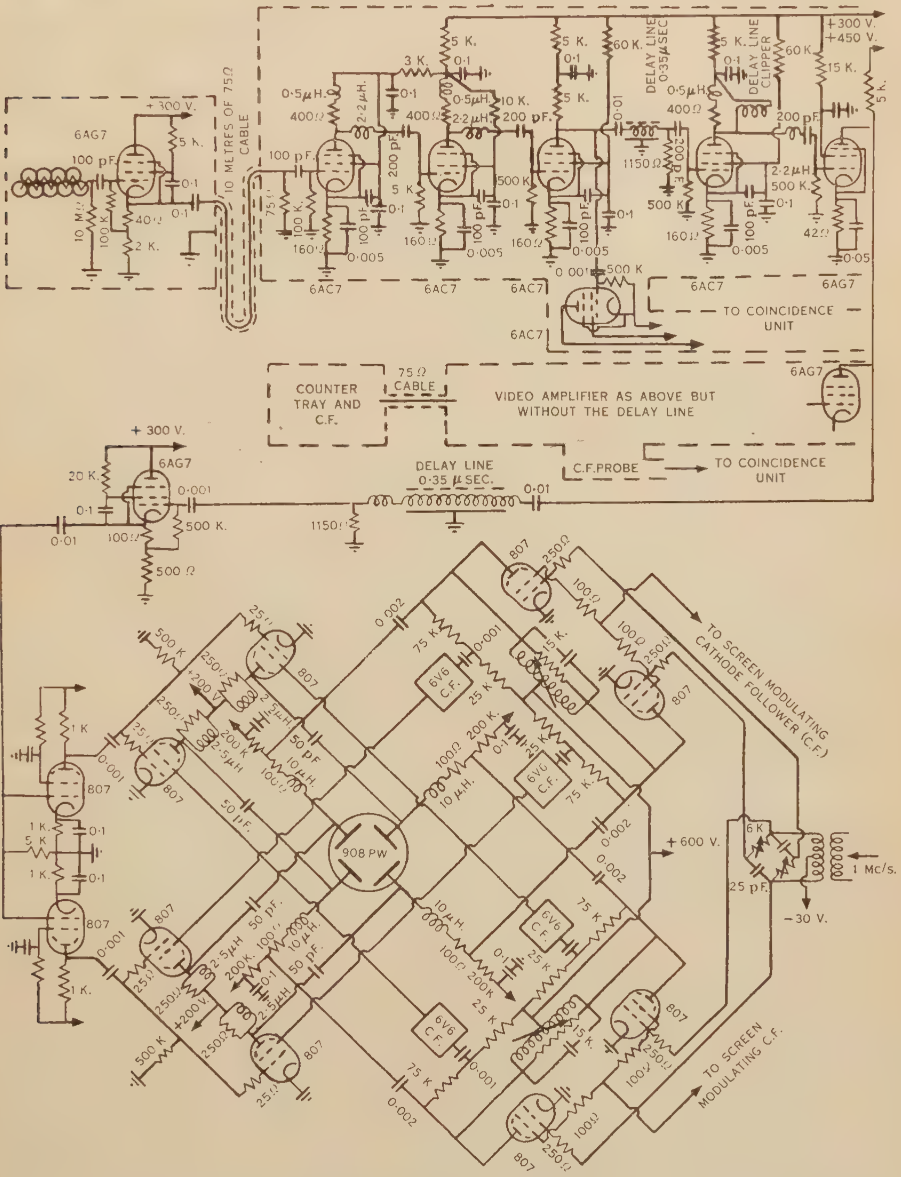

+300 V.
6AG7
100 pF.
5 K.
0·1
10 MΩ
100 K.
40Ω
2 K.
10 METRES OF 75Ω CABLE
75Ω
100 pF.
100 K.
0·5μH.
400Ω
2·2μH.
200 pF.
160Ω
100 pF.
0·005
3 K.
5 K.
0·1
10 K.
500 K.
60 K.
0·01
DELAY LINE 0·35μSEC.
1150Ω
200 pF.
500 K.
DELAY LINE CLIPPER
+300 V.
+450 V.
15 K.
5 K.
42Ω
0·05
0·001
6AC7
6AG7
TO COINCIDENCE UNIT
COUNTER TRAY AND C.F.
75Ω CABLE
VIDEO AMPLIFIER AS ABOVE BUT WITHOUT THE DELAY LINE
C.F.PROBE
DELAY LINE 0·35 μ SEC.
20 K.
0·01
100Ω
500 Ω
807
250Ω
25Ω
6V6 C.F.
75 K.
25 K.
200 K.
+200 V.
2·5μH.
10μH.
50 pF.
0·002
908 PW
1 K.
5 K.
+600 V.
6 K.
25 pF.
1 Mc/s.
−30 V.
TO SCREEN MODULATING CATHODE FOLLOWER (C.F.)
TO SCREEN MODULATING C.F.

Fig. 3.—Schematic diagram of video amplifier, radial deflection, and spiral deflection circuits.

This means that the voltage transients developed across the inductances are also proportional to the appropriate spiral signals, and the vector addition of the four transients results in radial deflection.

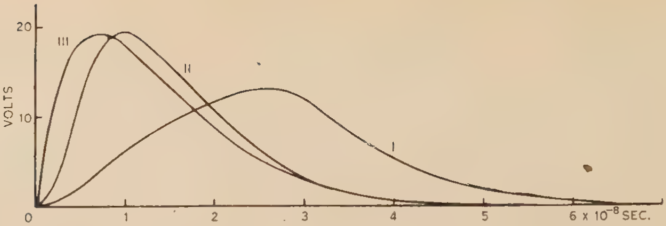


Fig. 4.—Transient waveforms at the plate of an 807 radial deflection tube. $L = 2 \cdot 5$ μH., $C = 20$ pF.

Case I. Tube cut off in $2 \cdot 5 \times 10^{-8}$ sec., i.e. $\alpha = 1200$ V./μsec.

Case II. Tube cut off in $0 \cdot 5 \times 10^{-8}$ sec., i.e. $\alpha = 6000$ V./μsec.

Case III. Tube cut off in zero time, i.e. $\alpha = \infty$.

All cases are for the peak of the screen modulation cycle (i.e. grid base* = 30 V. and plate current = 150 mA.).

The transient voltage appearing across an inductance L shunted by a condenser C and its critical damping resistor R, when a current i_0 flowing in it is instantaneously reduced to zero, has the well-known form

$$v = \frac{i_0 t}{C} e^{-t/2CR},$$

where t is the time measured from the instant of switching off and $2CR = \sqrt{LC}$.

When the switching off process takes place in a finite time according to the relation $i = i_0 - \alpha g_m t$, where α is the rate of decrease of grid voltage, it can be shown that

$$v = \alpha g_m L \left(1 - e^{-t/2CR} - \frac{t}{2CR} e^{-t/2CR}\right)$$

up to the time that the current reaches zero. The output voltage after this time can be found by assuming an equal and opposite input signal applied and combining the two outputs. The resulting transient wave forms for various values of α are shown in Figure 4. Figure 5 is a photograph of a typical spiral trace showing radial blips.

(c) *Video Amplifiers and Delay Lines*

In order to take advantage of the fast rising but small amplitude electron component of a Geiger counter pulse it is necessary to use wide band video amplifiers of moderate gain. When this electron pulse, which according to Kelly, Jordan, and Bell (1950) rises in about 10^{-8} sec., is used, virtually the only source of fluctuation in the reaction times of the Geiger counters is due to the time required by the electrons liberated by the ionizing event to drift from their place of formation to the neighbourhood of the central wire. Although the electron pulse rises in such a short time, satisfactory performance can be obtained with an amplifier and delay line having an overall rise time of 2 to 3×10^{-8} sec.

---

* The grid base, and therefore the instant of zero plate current relative to the beginning of the signal pulse, varies throughout the screen modulation cycle causing the four transients to have different shapes, but these differences are not serious provided the switch off is very rapid. This effect could probably be avoided by using suppressor modulated pentodes.

The video amplifiers used are of the conventional shunt-series compensated type (Fig. 3) having a bandwidth of 17 Mc/s. and a gain of 800. In one of them an extra delay line can be included to enable very short time intervals to be measured without the possibility of the two timing pulses overlapping. Delay line clippers reduce the pulses to 0·1 μsec. duration before they are mixed by the common anode method and fed into the main delay line.

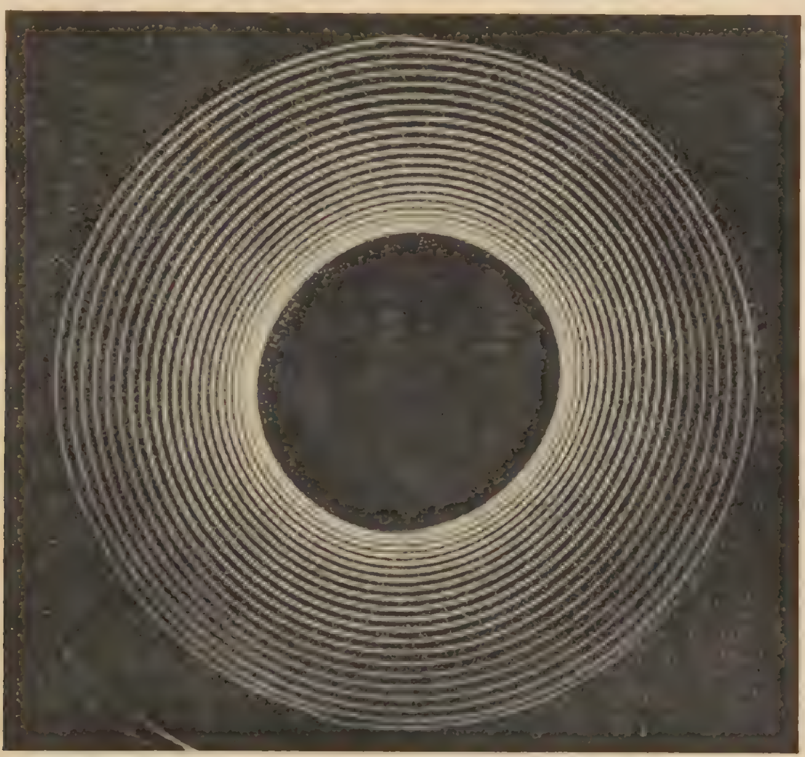


Fig. 5.—Section of photographic record showing radial blips on a spiral trace.

Both transmission delay lines are of the distributed constant type, having a delay of 0·35 μsec. and a characteristic impedance of 1150 ohms. The delay *versus* frequency characteristic has been compensated by using floating patches so that it is flat up to 20 Mc/s., and the amplitude characteristic has a pass band between the half-power points of 15 Mc/s. This has been achieved partly by following a suggestion by Kallmann (1946) that the line should be wound with several strands of 40 B & S wire in parallel to reduce skin effect, and partly by using bare wire wound in rectangular grooves in a polystyrene rod, thus making air and polystyrene the only dielectrics involved. A similar method of using bare wire was tried unsuccessfully by Kallmann. Figure 6 shows the details of the lines.

(*d*) *Spiral Generating and Other Circuits*

As the functions of radial deflection generation and spiral deflection voltage amplification are carried out in separate amplifiers, it is possible to use two pairs of 807 valves operated in class B with tuned plate loads as the spiral deflection amplifier. This allows considerable discrimination against harmonic distortion to be used and a consequent improvement in circle shape obtained, but conflicts with the requirement that the amplifier must be able to pass the 1 Mc/s. spiral signal of rapidly diminishing amplitude possessing very wide side bands. The difficulty is overcome by so distorting the spiral modulating pulse applied to the screens of the 807 valves that the modulation envelope produced has the desired shape.

The Rossi coincidence unit is fairly conventional in design except that steps are taken to increase the speed of response wherever possible. These include the use of high transconductance tubes, and the inclusion of amplifiers in each channel to reduce the fluctuations in the Geiger counter reaction times.

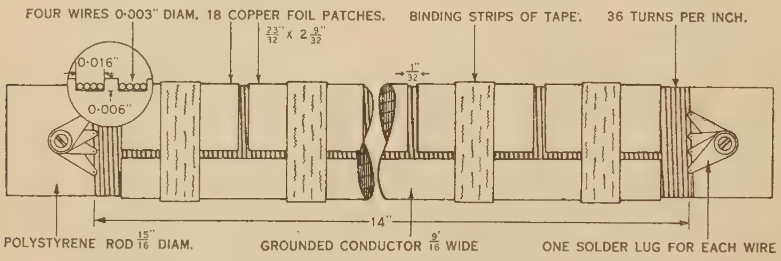

FOUR WIRES 0·003" DIAM.
18 COPPER FOIL PATCHES.
$\frac{23}{32}$" x $2\frac{9}{32}$"
BINDING STRIPS OF TAPE.
36 TURNS PER INCH.
0·016"
0·006"
$\frac{1}{32}$"
14"
POLYSTYRENE ROD $\frac{15}{16}$" DIAM.
GROUNDED CONDUCTOR $\frac{9}{16}$' WIDE
ONE SOLDER LUG FOR EACH WIRE

Fig. 6.—Distributed constant, equalized delay line.

The 10 kV. accelerating voltage and the Geiger counter high tension are obtained from radio-frequency high voltage power supplies. Power control is carried out in a separate unit incorporating thermal delays, which switches on the various parts of the circuit in the correct order and at the correct times. Since the equipment has to run continuously, various safety devices have been included in this unit. In addition to very light individual fuses for each unit of the circuit, a thyratron-operated main switch opens if any individual fuse blows, or if any important voltage fails. Automatic switch off also occurs if the rate of taking pictures becomes too high.

(*e*) *Accuracy*

The random errors inherent in the timing system were investigated by connecting pulses from the same Geiger counter to both timing channel inputs. These pulses should then be separated on the spiral by a constant time interval equal to the difference between the delays in the two channels, consisting mainly of the delay in the additional delay line in one of the channels. The distribution of 47 observations having a standard deviation of $5 \cdot 0 \times 10^{-9}$ sec. is included in Figure 1. Careful matching of tubes and components in the radial circuit, which has not yet been attempted, would probably improve this performance.

The absolute accuracy is more than adequate for the type of experiment intended, depending as it does on the stability of a quartz crystal oscillator and a knowledge of its frequency.

III. The Time of Flight Observations

The experiment suggested in the Introduction was carried out, but with one important refinement. Instead of making about 3000 observations with the trays close together to determine the time zero accurately, it is much easier to determine it in about 50 observations by applying the same pulse to both inputs and thus eliminating counter lags. This, however, involves the assumption that the two counter trays have identical average reaction times, but this objection can be overcome by interchanging the counters in the two trays half way through the observations with the trays separated.

The mean time of flight over $5\cdot45$ m. for 288 particles that could penetrate 10 cm. of lead was found to be $1\cdot9\pm0\cdot5\times10^{-8}$ sec. At the velocity of light this time would have been $1\cdot8\times10^{-8}$ sec.

There was a possibility that temperature effects in the delay line used to separate the two pulses on the spiral could have influenced this result. Therefore time zero determinations were made in the temperature range 15 to 36 °C., but no significant variation was found.

Extensive air showers were another source of error. In order to get a long path for single particles, the line joining the two timing trays was inclined at 35° to the horizontal, thus allowing showers, most of which come in near the vertical, to give false results. Shielding one tray with 10 cm. of lead reduced the shower rate. In addition a fourth counter tray (Fig. 7) was made to act as a shower monitor by causing a neon lamp to flash in view of the camera when it was discharged at the same time as the other three trays. By placing the middle tray of the three out of line, the shower monitor was shown to be only 80 per cent. efficient. This necessitated a correction to the time of flight for single particles, but it proved to be unimportant compared with the random error.

IV. The Shower Observations

The 92 shower photographs were analysed in the same way as the single particle ones, and the quantity analogous to the time of flight came to the unexpectedly large value of $3\cdot3\pm1\cdot1\times10^{-8}$ sec., a time lag associated with the shielded tray. On subtracting 1×10^{-8} sec., the expected time of flight over the vertical distance between the two trays, an unexplained $2\cdot3\pm1\cdot1\times10^{-8}$ sec. was left. This was only just significantly different from nothing, but a further experiment to see whether such a time lag really existed seemed to be warranted. It could perhaps be due to delayed particles in extensive air showers, or possibly to some internal cause.

Three trays were arranged in a horizontal plane (Fig. 8), the two timing trays being 4 m. apart, and 10 cm. of lead placed over tray A to determine whether the 2×10^{-8} sec. time lag was again obtained in tray A even though all three trays were at the same height. If it was still present and was some property of the showers, shifting the lead to tray *C* should reverse the sign of the time lag, since the time reading from channel *A* is always subtracted from that in channel *C*. Finally, placing the lead over tray *B* should give no average time lag in either timing channel.

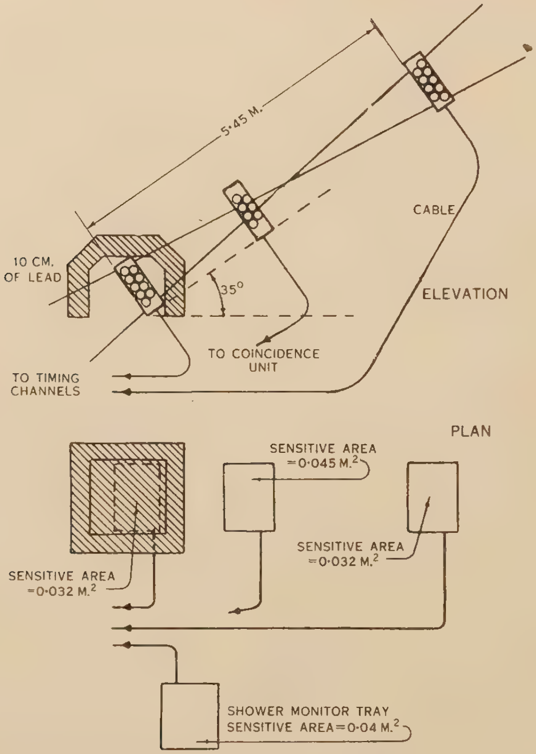

5·45 M.
10 CM. OF LEAD
35°
CABLE
ELEVATION
TO COINCIDENCE UNIT
TO TIMING CHANNELS
PLAN
SENSITIVE AREA =0·045 M.²
SENSITIVE AREA =0·032 M.²
SENSITIVE AREA =0·032 M.²
SHOWER MONITOR TRAY
SENSITIVE AREA=0·04 M.²

Fig. 7.—Time of flight counter tray arrangement (not to scale).

The experiment was carried out in this way, and as before the counters in trays *A* and *C* were interchanged half way through each run. The results are shown in Table 1.

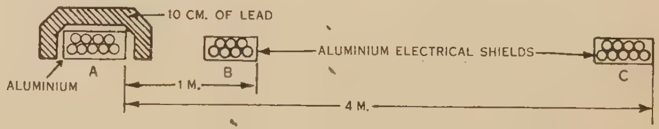

10 CM. OF LEAD
ALUMINIUM ELECTRICAL SHIELDS
ALUMINIUM
A
B
C
1 M.
4 M.

Fig. 8.—Shower experiment counter tray arrangement (not to scale).

The observed mean time lag associated with the shielded trays of $2{\cdot}1 \pm 0{\cdot}35 \times 10^{-8}$ sec., together with the fact that there is no lag significantly different from zero when the lead is over the middle tray, definitely establishes the existence of the phenomenon.

It may be noted that in 3 per cent. of the photographs a third pulse appeared between 1 and 2 μsec. after the first. These pulses with long delays could not have been accidental coincidences, but may have been due to negative ions in the counters, and little significance should be attached to them, especially in view of the fact that McCusker, Ritson, and Nevin (1950) found no such delayed particles in extensive air showers. They searched between 1·4 and 11·4 μsec. after detection of the showers.

TABLE 1

RESULTS OF SHOWER EXPERIMENT

Times expressed in sec. $\times 10^{-8}$

	Lead on *A*	Lead on *C*	Lead on *B*
1st Group			
No. of showers	105	104	—*
Mean time reading	39·51±0·68	43·77±0·73	
2nd Group			
No. of showers with counters interchanged	95	125	86
Mean time reading	38·56±0·71	42·80±0·62	39·81±0·54
Unweighted mean of 1st and 2nd groups	39·04±0·49	43·29±0·48	—*
Mean zero reading	41·03±0·07	41·03±0·07	40·55±0·35
Time lag associated with shielded tray	1·99±0·50	2·26±0·49	0·74±0·65†

Mean time lag associated with the shielded trays *A* and *C* for the 429 showers : 2·1±0·35.

* Counter failure forced the termination of the measurements with lead on tray *B*, but since the difference between the average reaction times of the two counter groups was known from the previous observations the appropriate zero reading could be deduced. It is

$$41\cdot03-\tfrac{1}{2}\left[\frac{(39\cdot51-38\cdot56)+(43\cdot77-42\cdot80)}{2}\right]=40\cdot55\pm0\cdot35.$$

† This is a time lag associated with tray *A*, but it is not significantly different from zero.

V. INTERPRETATION OF THE TIME LAG

It is conceivable that an extensive air shower, if frozen in space at a particular instant, would have an appreciable vertical thickness. If the fast particles which penetrate the lead always arrive first, a time lag should be associated with the unshielded tray, but this is contrary to experiment. If the small number of penetrating particles are distributed evenly throughout the thickness of the shower, a time lag would be associated with the shielded tray, because its average position of firing would be deeper in the shower than for the unshielded tray exposed to the full shower radiation. The first possibility is, however, not excluded, because of the presence of a much larger effect which must now be considered.

Only a small fraction of the shower particles penetrate to the shielded tray, and consequently on most occasions the shielded tray may be traversed by only one particle, while the unshielded timing tray is hit by several particles. This would result in the smallest of several randomly chosen counter reaction times being accepted for the unshielded tray, and subtracted from one randomly chosen reaction time from the shielded tray. The average reaction time from the shielded tray would be the larger, giving in effect a time lag associated with that tray.

It is shown in Appendix I that 11 per cent. of the shower radiation must be assumed to penetrate the 10 cm. lead shielding, in order to explain the observed shower rate of $1 \cdot 54 \pm 0 \cdot 14$ per hour. This means that the penetrating radiation cannot be the 1 per cent. meson component of extensive air showers, but must be the more penetrating electrons and photons of the air showers, or the photons that survive the burnt out showers started by them in the lead, as pointed out by Greisen (1949).

Using the proportion of penetrating radiation thus found, the theory of the internal effect is worked out in Appendix I, and the calculation carried far enough to include the effect of 75 per cent. of the recorded showers. For this fraction of the showers the time lag associated with the shielded tray is found to be $2 \cdot 09 \times 10^{-8}$ sec.

The good agreement between the calculated time lag and the experimental value $2 \cdot 1 \pm 0 \cdot 35 \times 10^{-8}$ sec. is fortuitous, because no account has been taken of the fact that the counters were overlapped, giving an equivalent diameter, for particles incident vertically, of half the actual counter diameter ; but it is almost certain that scattering greatly reduced the effectiveness of this overlapping. Non-vertical arrival of the showers would contribute, but scattering in the lead and firing of the shielded counters by photons would be more important, increasing the effective diameter of the overlapped counters in the shielded tray and increasing the time lag associated with that tray. These two compensating effects provide the main uncertainty in the interpretation of the results.

If unlapped counters had been used, it may have been possible to make some deduction about shower thickness by comparing the calculated and observed time lags. As it is, we are justified in stating only that the internal effects appear to be capable of explaining the observed time lag, and that consequently the majority of particles in a shower probably arrive at the apparatus within a time less than 10^{-8} sec. This view is supported by the results of an approximate calculation by Mr. K. C. Hines of this Laboratory, giving the maximum vertical thickness of a frozen shower near its core, if the spread is accumulated only in the last radiation length and particles with energies down to 1 MeV. are considered. The resulting maximum time interval for the arrival of the shower particles is about 10^{-9} sec. Mr. Hines has also pointed out the possibility of measuring with this apparatus, a time difference between the arrival of the core of a shower and its outer edge at the surface of the Earth, thus providing a check on shower diffusion theory.

VI. ACKNOWLEDGMENTS

The writer is greatly indebted to Mr. J. S. Grice who helped with the construction and development of the apparatus, and to Mr. G. F. Mauldon for help with the observations. He wishes to thank Dr. H. D. Rathgeber for helpful suggestions and Professor L. H. Martin for his continued interest in this work. Part of the cost was met from funds provided by C.S.I.R.O.

VII. REFERENCES

GREISEN, K. (1949).—*Phys. Rev.* **75** : 1071.
HARTOG, H. DEN, MULLER, F. A., and VERSTER, N. F. (1947).—*Physica* **13** : 251.
KALLMANN, H. E. (1946).—*Proc. Inst. Radio Engrs. N.Y.* **34** : 646.
KELLY, G. G., JORDAN, W. H., and BELL, P. R. (1950).—*Rev. Sci. Instrum.* **21** : 330.
MCCUSKER, C. B. A., RITSON, D. M., and NEVIN, T. E. (1950).—*Nature* **166** : 400.
WILKINSON, D. H. (1950).—"Ionization Chambers and Counters." (Cambridge Univ. Press.)

APPENDIX I

To calculate from the observed shower rate of $1\cdot54\pm0\cdot14$ per hour, the fraction ε of the shower radiation penetrating the 10 cm. lead shielding, consider the well-known expression for the shower rate

$$N=K\int_0^\infty \Delta^{-\gamma}(1-e^{-S_1\varepsilon\Delta})(1-e^{-S_1\Delta})(1-e^{-S_2\Delta})d\Delta, \quad \ldots\ldots \quad (1)$$

where Δ is the shower density in particles per m.^{2},

S_1 and S_2 are the areas of the timing trays and the triple coincidence tray respectively,

$K\Delta^{-\gamma}$ is the differential spectrum of shower densities with $\gamma=2\cdot5$,

K was found to be $1\cdot48\times10^3$ for this counter arrangement by measuring the rate without any lead.

The other factors are the probabilities of discharge of the various trays.

Graphical solution gives $\varepsilon=11$ per cent.

The probability that a tray will be hit by r particles, where r has the values 1, 2, 3.... during a shower of density Δ, must now be determined and integrated over all shower densities. For a tray of area S the average number of particles crossing it per shower of density Δ is $S\Delta$. The number of particles crossing it in a particular shower must obey the Poisson distribution. Hence the probability for r particles crossing the tray is

$$P_r=\frac{e^{-S\Delta}(S\Delta)^r}{r!}, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

and

$$\sum_{r=1}^{r=\infty} P_r=1-e^{-S\Delta}$$

the probability for discharge of the tray by any number of particles. Hence from equations (**1**) and (**2**) the fraction of all the recorded showers with r particles through the shielded and n through the unshielded tray is

$$\varphi_{r,n}=\frac{\int_0^\infty \Delta^{-\gamma}\frac{e^{-S_1\varepsilon\Delta}(S_1\varepsilon\Delta)^r}{r!}\cdot\frac{e^{-S_1\Delta}(S_1\Delta)^n}{n!}(1-e^{-S_2\Delta})d\Delta}{\int_0^\infty \Delta^{-\gamma}(1-e^{-S_1\varepsilon\Delta})(1-e^{-S_1\Delta})(1-e^{-S_2\Delta})d\Delta} \quad \ldots \quad (3)$$

Table 2 shows the values of $\varphi_{r,n}$ obtained graphically.

TABLE 2

VALUES OF $\varphi_{r,n}$

$\varphi_{r,n}$		n					
		1	2	3	4	5	6
r	1	0·345	0·159	0·090	0·0508	0·0352	0·0245
	2	0·018	0·011	0·009			
	3	0·0011	0·0013	0·0016			

Then if the corresponding relative counter time lag distributions are $Y_{r,n}(\tau)$ and their mean values $\mu_{r,n}$ the mean of the distribution, including all showers detected, is given by the weighted mean

$$M=\sum_{r=1}^{r=\infty}\sum_{n=1}^{n=\infty}\varphi_{r,n}\mu_{r,n}\Big/\sum_{r=1}^{r=\infty}\sum_{n=1}^{n=\infty}\varphi_{r,n}. \qquad \ldots\ldots\ldots\ldots\ldots \quad (4)$$

In order to find the $\mu_{r,n}$ the absolute and relative counter time lag distributions must be calculated. The calculation of the counter absolute time lag, or reaction time distribution, requires a knowledge of the mobility of electrons at the electric fields X and pressures p met with in Geiger counters. Wilkinson (1950) gives the electron velocity $v=\alpha X/p$, where $\alpha=1{\cdot}6\times10^6$ from the work of den Hartog, Muller, and Verster (1947), but Wilkinson points out that this relation is obtained by extrapolation based on slender evidence.

For a cylindrical counter with wire radius a and cylinder radius b operating at a voltage V, it is easily shown that the time taken by an electron to drift from a point distant x from the centre of the wire to the surface of the wire is

$$t=\beta(x^2-a^2),$$

where $$\beta=\frac{p\log(b/a)}{2\alpha V},$$

and therefore $$x=\beta^{-\frac{1}{2}}(t+\beta a^2)^{\frac{1}{2}}.$$

Now, in a given time, let m particles from a uniform parallel beam incident normally on a plane containing the counter wire fall on a strip of the plane parallel to the wire, of unit width, and as long as the counter. Then the number of particles giving time lags $>t$ is

$$N=m(b-x),$$

which gives the integral distribution of time lags. The differential distribution is then

$$-dN/dt=(m/2\beta^{\frac{1}{2}})(t+\beta a^2)^{-\frac{1}{2}},$$

and the distribution normalized to unit area under the curve is

$$F(t)=\nu(t+\sigma)^{-\frac{1}{2}}, \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

where
$$\sigma=\beta a^2,$$
$$\nu=1/2\beta^{\frac{1}{2}}(b-a),$$
and t goes from 0 to t_{max}, where t_{max} is the reaction time when $x=b$.

$F(t)$ is the absolute time lag distribution for one particle at a time striking the counter, but the distribution of the shortest time lag when r particles arrive simultaneously is required. For the case when $r=2$, consider N pairs of simultaneous hits on a counter.

The number of pairs containing one lag between t_0 and t_0+dt_0 is
$$N.F(t_0)dt_0,$$
and these pairs have their second member distributed according to
$$N.F(t_0)dt_0 \cdot F(t)dt. \quad \dots\dots\dots\dots\dots\dots\dots \quad (6)$$
Introduce a new variable τ to represent the shorter of the two time lags of a pair.

When t_0 is the shorter $t_0=\tau$ and the contribution to the differential distribution of τ is
$$Z_1(\tau)d\tau=N\int_{t=\tau}^{t=t_{max}} F(\tau)d\tau F(t)dt.$$
When t is the shorter $t=\tau$ and the contribution is
$$Z_2(\tau)d\tau=N\int_{t_0=\tau}^{t_0=t_{max}} F(t_0)dt_0F(\tau)d\tau.$$
Therefore the complete distribution for $r=2$ is
$$\underset{r=2}{Z}(\tau)=2N.F(\tau)\int_{\tau}^{t_{max}} F(t)dt, \quad \dots\dots\dots\dots\dots\dots\dots \quad (7)$$
which, on integration, becomes
$$\underset{r=2}{Z}(\tau)=4N\nu^2\left[\left(\frac{t_{max}+\sigma}{\tau+\sigma}\right)^{\frac{1}{2}}-1\right]=N.W(\tau). \quad \dots\dots \quad (8)$$
Similarly the corresponding expression to (6) when $r=3$ is
$$N.F(t_0)dt_0.W(t)dt,$$
and the resulting distribution of the shortest of three randomly chosen time lags is
$$\underset{r=3}{Z}(\tau)=N.F(\tau)\int_{\tau}^{t_{max}} W(t)dt+N.W(\tau)\int_{\tau}^{t_{max}} F(t)dt, \quad \dots\dots\dots \quad (9)$$
which, on integration, becomes
$$\underset{r=3}{Z}(\tau)=4N\nu^3\left[\frac{3t_{max}+4\sigma+\tau}{(\tau+\sigma)^{\frac{1}{2}}}-6(t_{max}+\sigma)^{\frac{1}{2}}+2(\tau+\sigma)^{\frac{1}{2}}\right], \quad \dots\dots \quad (10)$$
$$=N.\varphi(\tau).$$
These distributions for $r=1$, 2, 3 are plotted in Figure 9.

The relative counter time lag distributions must now be established by finding the distributions of the differences between pairs of absolute lags chosen at random from the same absolute distribution or different absolute distributions.

Suppose $Y_{r,n}(\tau)$ is the relative distribution in which one lag is chosen from the absolute distribution with $r=r$ and the other from the distribution for $r=n$.

To find $Y_{1,1}(\tau)$ consider all the pairs of absolute lags out of N chosen from the counters A and B in which the lag in A lies between t_0 and t_0+dt_0.

Their number is $N.F(t_0)dt_0$.

The corresponding lags in B are distributed according to

$$N.F(t_0)dt_0.F(t).$$

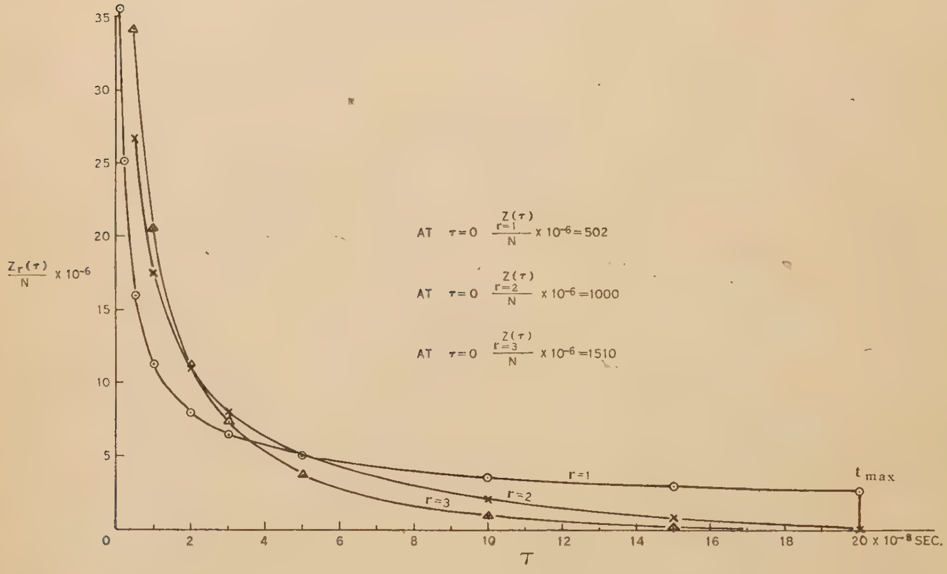


Fig. 9.—Distributions of absolute time lags.

Put $\tau=t-t_0$, the difference between the two absolute lags, and integrate with respect to t_0, giving

$$\frac{Y_{1,1}(\tau)}{N}=\int F(t_0)F(t_0+\tau)dt_0 \qquad \text{between the limits 0 to } t_{max}-\tau \text{ for } \tau \text{ positive and } -\tau \text{ to } t_{max} \text{ for } \tau \text{ negative.}$$

$$=\nu^2 \log \frac{\left[\sigma+t_{max}\mp\frac{\tau}{2}+\{(\sigma+t_{max})^2\mp\tau(\sigma+t_{max})\}^{\frac{1}{2}}\right]}{\left[\sigma\pm\frac{\tau}{2}+(\sigma^2\pm\tau\sigma)^{\frac{1}{2}}\right]}, \quad \ldots\ldots\ldots \quad (11)$$

where the upper of the alternative signs are to be taken with τ positive, and the lower ones with τ negative, giving a symmetrical distribution, as is to be expected. The mean value of τ for this distribution, $\mu_{1,1}$, is therefore zero.

Similarly,

$$Y_{1,2}(\tau)=N\int F(t)W(t+\tau)dt \qquad \text{between the limits 0 to } t_{max}-\tau \text{ for } \tau \text{ positive and } -\tau \text{ to } t_{max} \text{ for } \tau \text{ negative.}$$

On integration this becomes, for τ positive,

$$\left.\begin{array}{l} \dfrac{Y_{1,2}(\tau)}{N}=4\nu(t_{max}+\sigma)^{\frac{1}{2}}\dfrac{Y_{1,1}(\tau)}{N}-8\nu^3[(t_{max}+\sigma-\tau)^{\frac{1}{2}}-\sigma^{\frac{1}{2}}] \\ \text{and for } \tau \text{ negative,} \\ \dfrac{Y_{1,2}(\tau)}{N}=4\nu(t_{max}+\sigma)^{\frac{1}{2}}\dfrac{Y_{1,1}(\tau)}{N}-8\nu^3[(t_{max}+\sigma)^{\frac{1}{2}}-(\sigma-\tau)^{\frac{1}{2}}] \end{array}\right\}, \quad \ldots . \ (12)$$

which is a skew distribution.

The mean value of τ for this distribution is given by

$$\mu_{1,2}=\frac{1}{N}\int_{-t_{max}}^{t_{max}}\tau Y_{1,2}(\tau)d\tau.$$

The first term is zero by symmetry and the other term gives

$$\mu_{1,2}=-8\nu^3\left[\frac{2}{3}t_{max}(\sigma+t_{max})^{3/2}-\frac{2}{3}t_{max}\sigma^{3/2}-\frac{t^2_{max}\,\sigma^{\frac{1}{2}}}{2}-\frac{t^2_{max}}{2}(t_{max}+\sigma)^{\frac{1}{2}}\right] \quad \ldots\ldots\ldots\ldots (13)$$

$$=-3\cdot32\times10^{-8}\text{ sec.}$$

on substituting $t_{max}=2\times10^{-7}$ sec. and

$\sigma=5\times10^{-12}$, obtained by setting

$p=130$ mm. of Hg,
$b=1\cdot0$ cm.,
$a=0\cdot005$ cm.,
$V=1080$ V.,
$\alpha=1\cdot59\times10^{6}$ cm. sec.$^{-1}$ per X/p.

Following the argument used for the establishment of $Y_{1,1}(\tau)$ and $Y_{1,2}(\tau)$ it is seen that

$$Y_{1,3}(\tau)=N\int F(t)\varphi(t+\tau)dt, \qquad \text{for the range } \begin{array}{l} 0\leqslant t\leqslant t_{max}-\tau \text{ for } \tau \text{ positive,} \\ -\tau\leqslant t\leqslant t_{max} \text{ for } \tau \text{ negative.} \end{array}$$

On integrating this becomes

$$\frac{Y_{1,3}(\tau)}{N}=-12\nu^3(t_{max}+\sigma)\frac{Y_{1,1}(\tau)}{N}+6\nu(t_{max}+\sigma)^{\frac{1}{2}}\frac{Y_{1,2}(\tau)}{N}+6\nu^2\frac{\tau Y_{1,1}(\tau)}{N}$$
$$+12\nu^4[\{(t_{max}+\sigma)^2\mp\tau(t_{max}+\sigma)\}^{\frac{1}{2}}-(\sigma^2\pm\sigma\tau)^{\frac{1}{2}}], \quad \ldots\ldots\ldots (14)$$

where the upper alternative signs are to be taken with τ positive and the lower ones with τ negative.

This skew distribution together with $Y_{1,1}(\tau)/N$ and $Y_{1,2}(\tau)/N$ are plotted in Figure 10.

As in the case of $\mu_{1,2}$ the mean value $\mu_{1,3}$ is given by

$$\mu_{1,3}=\frac{1}{N}\int_{-t_{max}}^{t_{max}}\tau Y_{1,3}(\tau)d\tau$$
$$=6\nu(t_{max}+\sigma)^{\frac{1}{2}}\mu_{1,2}+\frac{6\nu^2}{N}\int_{-t_{max}}^{t_{max}}\tau^2 Y_{1,1}(\tau)d\tau,$$

since the other two terms are symmetrical in τ and therefore vanish.

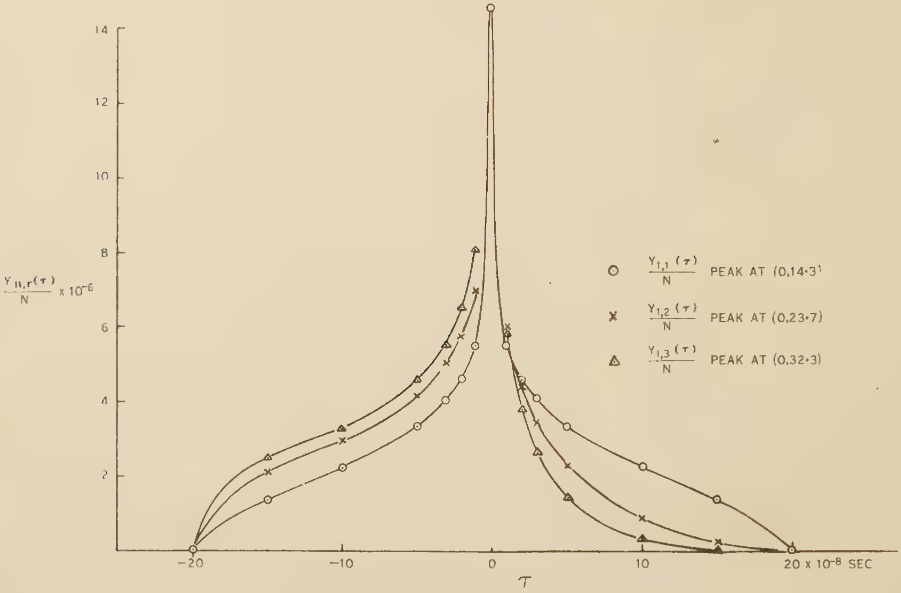


Fig. 10.—Distributions of relative time lags.

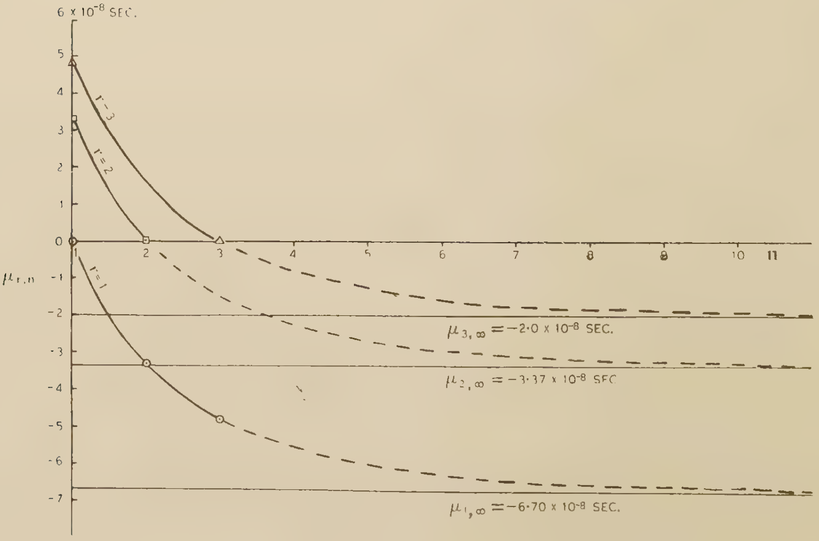


Fig. 11.—Mean values of τ, i.e. $\mu_{r,n}$, for the relative distributions $Y_{r,n}(\tau)$.

The remaining integral can be evaluated graphically, giving

$$\mu_{1,3} = -4 \cdot 79 \times 10^{-8} \text{ sec.}$$

It can readily be seen that the limiting value $\mu_{1,\infty}$ is simply plus or minus the mean value of τ for the absolute distribution $F(\tau)$. This follows from the fact that when an infinite number of particles hit one tray the probability for zero time lag in that tray will be very high. The negative sign arises when the lag in the other tray is subtracted from that in the last mentioned.

$$\mu_{1,\infty} = -6 \cdot 70 \times 10^{-8} \text{ sec.},$$

and similarly it can be shown that

$$\mu_{2,\infty} = -3 \cdot 37 \times 10^{-8} \text{ sec.},$$

and $$\mu_{3,\infty} = -2 \cdot 0 \times 10^{-8} \text{ sec.}$$

The values of $\mu_{r,n}$ are plotted in Figure 11. Since the $\mu_{r,\infty}$ values define the tangents to the curves at $n = \infty$, extrapolation can be carried out with sufficient accuracy for the present purpose. This process is aided by the discovery that the curves all have nearly the same shape, being merely displaced upwards towards a limiting position as r increases.

The sum of the 12 calculated fractions $\varphi_{r,n}$ in Table 2 accounts for 75 per cent. of the recorded showers. From equation (4) the mean of the composite relative distribution for this fraction of the recorded showers is found to be

$$M = -2 \cdot 09 \times 10^{-8} \text{ sec.}$$

OPERATIONAL IMPEDANCE MATRICES OF n-PHASE PARTIALLY-SYMMETRICAL MACHINES

By D. W. C. Shen*

[*Manuscript received July* 23, 1951]

Summary

By an affine transformation of space with origin fixed, the hyper-space associated with an n-phase partially-symmetrical machine having multi-winding in the field system is partitioned into two orthogonal sub-spaces of $n-2$ and 2 dimensions respectively. The armature reaction and the saliency effect of the machine are found to depend only upon the interaction of two rotating axes on the 2-dimensional plane. The complex stationary axes in the $n-2$ dimensional sub-space do not react on the field system and hence are named the zero armature-reaction axes. This supports a conclusion that, barring single-phase machines, the transient and steady-state performance characteristics of machines of any number of phases under balanced load conditions are essentially the same.

For transient solutions, the cases of suddenly-applied voltages and sudden short-circuits are considered. The correlation between the machine parameters as obtained in the present analysis with those obtained by the two-reaction method is established.

Notation

n=number of phases,

$i_a, i_b, \ldots i_n$=instantaneous value of armature currents in phases $a, b, c, \ldots n$,

$v_a, v_b, \ldots v_n$=instantaneous value of armature terminal potentials in phases $a, b, c, \ldots n$,

i_{df} =instantaneous value of direct-axis field current,

i_{qf} =instantaneous value of quadrature-axis field current,

E_{df} =D.C. potential of direct-axis field circuit,

E_{qf} =D.C. potential of quadrature-axis field circuit,

ω =synchronous angular velocity,

p $=d/dt$=the time differential operator,

k =one of the nth roots of unity,

θ_0 =the phase displacement of armature winding axes,

m =number of circuits in the direct axis of the field side,

m' =number of circuits in the quadrature axis of the field side,

R_a =resistance of armature, or stator, per phase,

L_{aa} =constant part of the self-inductance of armature, or stator, per phase,

M'_a=maximum value of the constant part of the mutual inductance between armature phases,

L'_{a2}=maximum value of the second harmonic of the self-inductance of armature per phase; it is also equal to the maximum value of the second harmonic of the mutual inductance between phases,

---

* Electrical Engineering Department, University of Adelaide.

$L_{a2} = \frac{n}{2}L'_{a2}$,

$L_a = L_{aa} - M_a'$ = zero-sequence inductance of armature,

$L_{a0} = L_a + \frac{n}{2}M_a' = L_{aa} + \left(\frac{n}{2}-1\right)M_a'$ = constant part of the synchronous self-inductance of armature, or stator, per phase; it is the constant self-inductance part of one phase plus the mutual effects of the other phases,

$L_{ad} = L_{a0} + L_{a2}$ = direct-axis synchronous inductance,

$L_{aq} = L_{a0} - L_{a2}$ = quadrature-axis synchronous inductance,

R_{df} = resistance of direct-axis field windings,

L_{df} = self-inductance of direct-axis field windings,

M_{df} = mutual inductance between direct-axis field windings,

M_d = maximum value of the mutual inductance between one phase of the armature and the direct-axis field windings,

R_{qf} = resistance of quadrature-axis field windings,

L_{qf} = self-inductance of quadrature-axis field windings,

M_{qf} = mutual inductance between quadrature-axis field windings,

M_q = maximum value of the mutual inductance between one armature phase and the quadrature-axis field windings.

I. INTRODUCTION

Based on the Hamilton-Cayley identity, the author (1948) has previously shown that the existing sets of coordinates used in the analysis of rotating machinery and power networks are simply related by rotation of space. Kron's (1935, 1950) papers on tensor analysis employed analogous ideas as that of Sah (1936), namely space vectors and dyadics, or tensors, in a much more extensive and ambitious programme, in which all types of rotating machinery were studied from a uniform point of view. Of the several different features of tensor analysis, the most widely known are the index notation and the technique of transforming from one coordinate system to another. It is these two features for which so much is claimed in the solution of electrical-circuit problems. The transformation matrix is perhaps the centre of study and an attempt is made here to point out explicitly what relations it has with the subject in general.

The equation of motion in the holonomical dynamical system is given by

$$e_k = L_{mk}\frac{di^m}{dt} + [m\,n,k]i^m i^n + R_{mk}i^m, \qquad \dots\dots\dots\dots \quad (1)$$

where L_{mk} = a fundamental or metric tensor,

i^m = velocity or current vector,

R_{mk} = friction or resistance vector,

e_k = applied force or voltage vector,

$[m\,n, k]$ = Christoffel symbol of the first kind; it is a geometric object of rank 3.

The occasion may arise when it is preferred to work with a set of generalized velocities other than the set at hand, but since both sets describe the same system, they must be related by a group of equations, such as

$$i^{\alpha}=A_m{}^{\alpha}i^m, \qquad \dots\dots\dots\dots\dots\dots\dots\dots \quad (2)$$

where the index α indicates the new system and the index m indicates the old system. Here a transformation is carried out and the transformation matrix, $A_m{}^{\alpha}$, transforms the old system into a new one, each having the same number of degrees of freedom, and the matrix is non-singular so that inverse transformation is always possible. Very often the equations represented by (2) are not integrable; in such a case the new generalized velocities are called quasi-coordinates (Kron 1936) and they are not true Lagrangian coordinates.

It is noteworthy that, in the case of electrical machinery in which the coordinates are taken to be charges in the different conductors together with the angular displacement of the rotor with respect to the stator, the inductance tensor is a function of θ alone. As in quasi-holonomic transformation of such coordinates, θ is usually not subjected to any transformation; the equation of motion in the new coordinates can be simplified and put into the following form (Kron 1942):

$$e_{\gamma}=L_{\gamma\alpha}\frac{di^{\alpha}}{dt}+[\alpha\,\beta,\gamma]i^{\alpha}i^{\beta}+S_{\alpha\beta\gamma}i^{\alpha}i^{\beta}+R_{\gamma\alpha}i^{\alpha}, \qquad \dots\dots \quad (3)$$

where $S_{\alpha\beta\gamma}$ is known as the torsion tensor which gives the arrangement of inductances due to rotation of the conductors. The usual effect of changing to quasi-coordinates is to increase the complexity of the equations; therefore, in advanced dynamics, quasi-coordinates would not be used for choice, but would be used for non-holonomic systems (Kron 1934). The quasi-coordinates used in electrical machines, however, are of a very special kind; they allow the equation of performance to emerge in the simplest form. It has been shown by Kron (1942) that in most standard electrical machines with stationary axes, the holonomic Christoffel symbol $[\alpha\,\beta,\gamma]$ vanishes. The direct and quadrature currents of the two-reaction theory in synchronous machines are quasi-coordinates because equation (2) is not integrable if i were replaced by dq/dt. The use of quasi-coordinates simplifies many machine problems and its advantages have recently been discussed by Gibbs (1950). In electrical machines, we are interested in the currents and are not concerned with the actual coordinates themselves. This is the main reason why the solutions of electrical machinery problems can be obtained much more easily than those of other dynamical problems.

Park (1929) and Laudner (1936) applied the above transformation to three-phase synchronous machines. Ku (1937) later extended the two-reaction theory to multi-phase machines. In short, the original phase coils form a mutually-orthogonal system electrically, while the new coordinates form a mutually-orthogonal system magnetically.

In this paper, the field windings of a generalized partially-symmetrical machine are considered as stationary, but the armature coils are transformed to complex axes, $n-2$ of which are stationary, while the remaining two are rotating.

This partition enables the stationary axes to become independent of the field axes, as the field axes are along the direct and quadrature axes magnetically. The transformation is unitary, characterized by the property that the reciprocal of the matrix (or tensor) is equal to the transpose of the conjugate (or the conjugate of the transpose). The same transformation can be used to analyse the behaviour of asynchronous machines with asymmetrical rotors.

II. Assumptions Regarding Machine Structure

In the transient analysis of A.C. machines, the problems are readily grouped in three classes. In the first, the circuits on both sides of the air-gap are symmetrical; in the second, the circuit on but one side of the air-gap is symmetrical; and in the third, the circuits on neither side of the air-gap are symmetrical. We shall study the operational impedance matrices of the most general type of partially-symmetrical machines; that is, a machine having any number of phases greater than one in the armature circuit and also having any number of windings in the field system. The field structure of such a machine may be completely symmetrical or may depart from complete symmetry due to saliency of poles.

To consider an actual machine with its every little detail of electric and magnetic structure taken care of by rigorous mathematics is well-nigh impossible, and in all theoretical consideration of such problems, some assumptions which are approximately true in practice have always been made (Park 1928). In the present analysis, the following will be taken to hold:

(i) Hysteresis and saturation in the iron are neglected.

(ii) The effects of slots on flux distribution are neglected.

(iii) Resistances of windings are constant for all positions between winding axes.

(iv) Inductances of windings, self or mutual, are constant for a given relative position of the windings.

(v) The mutual inductance between a stationary and a rotating winding varies with the cosine of the angle between the winding axes.

(vi) The stator (considered as armature) windings are exactly alike except for a phase displacement of $360/n$ degrees of the axes. The self-inductance of each as well as the mutual inductance between any two are equal to a constant plus a term that varies second-harmonically as the angular position of the rotor.

(vii) In general, there may be m circuits in the direct axis and m' circuits in the quadrature axis of the field side.

III. Differential Equation

The transformation from one coordinate system to another, while a major part of tensor analysis, is also fundamental to the theory of matrix algebra (Vowels 1944). It is true that the index notation has its merits, for example the formulation of the Christoffel symbols is greatly facilitated by the index notation. However, the distinct advantages of index notation over other forms in the formulation of n-way tensors ($n>2$) disappear in the manipulation of one-way or two-way forms of electrical networks. In fact, Kron (1942) expresses the tensor transformation formulae conveniently in matrix forms, but emphasizes that the interpretation is based on tensor concepts.

The equations of performance for a partially-symmetrical machine may be written in matrix notation as follows (Sah 1939):

$$\mathbf{v}_a=\mathbf{Z}_{aa}\mathbf{I}_a+p\mathbf{M}_{af}\mathbf{I}_f, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (4a)$$

$$\mathbf{E}_f=p\mathbf{M}_{fa}\mathbf{I}_a+\mathbf{Z}_{ff}\mathbf{I}_f, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (4b)$$

where $p=\frac{d}{dt}$, $\mathbf{v}_a=\{v_a, v_b, \ldots . v_n\}$ is the voltage vector whose components are the instantaneous armature terminal potentials in phases, a, b, $c, \ldots . n$, and $\mathbf{I}_a=\{i_a, i_b, \ldots . i_n\}$ is the corresponding armature current vector. (Curly brackets are used here to denote column-matrices written in a row.)

$\mathbf{E}_f$ is the field voltage vector which may be partitioned into two groups such as:

$$\mathbf{E}_f=\{E_{df1}E_{df2}\ldots . E_{dfm} \quad E_{qf1}E_{qf2}\ldots . E_{qfm'}\},$$

E_{df} being the D.C. potentials of direct-axis field circuits, and E_{qf} the D.C. potentials of quadrature-axis field circuits.

$\mathbf{I}_f$ is the field current vector which may be similarly partitioned, that is

$$\mathbf{I}_f=\{i_{df1}i_{df2}\ldots . i_{dfm} \quad i_{qf1}i_{qf2}\ldots . i_{qfm'}\},$$

i_{df} and i_{qf} being the instantaneous values of direct- and quadrature-axis field currents respectively.

$\mathbf{Z}_{aa}$ and $\mathbf{Z}_{ff}$ are the instantaneous self-impedance matrices of the armature and the field respectively.

$\mathbf{M}_{af}$ and $\mathbf{M}_{fa}$ are the mutual inductance matrix and its transpose of the rotor with respect to the stator.

The self-impedance matrix, $\mathbf{Z}_{aa}$, of an n-phase partially-symmetrical machine may be separated into two parts

$$\mathbf{Z}_{aa}=\mathbf{Z}_{aa0}+\mathbf{Z}_{aa2}.$$

$\mathbf{Z}_{aa0}$ is the constant part of the matrix, whereas $\mathbf{Z}_{aa2}$ varies second-harmonically with the rotor position and therefore characterizes the saliency feature of the machine.

$\mathbf{Z}_{aa0}$ will be the usual form of a symmetrical network, that is

$$\mathbf{Z}_{aa0}=\begin{bmatrix} R_a+pL_{aa} & pM_a'\cos\theta_0 & \ldots\ldots & pM_a'\cos(n-1)\theta_0 \\ pM_a'\cos\theta_0 & R_a+pL_{aa} & \ldots\ldots & pM_a'\cos(n-2)\theta_0 \\ \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots\ldots & \ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots\ldots & \ldots\ldots\ldots\ldots\ldots \\ pM_a'\cos\theta_0 & pM_a'\cos 2\theta_0 & \ldots\ldots & R_a+pL_{aa} \end{bmatrix} \quad \ldots (5a)$$

In equation (**5a**), R_a is the resistance of armature or stator, per phase ; L_{aa} is the constant part of the self-inductance of armature or stator, per phase ; M_a' is the maximum value of the constant part of the mutual inductance between armature phases ; and $\theta_0=360°/n$ is the phase displacement of armature-winding axes. It should be noted that M_a', though a fictitious value, is the mutual inductance between two armature phases if their axes were to be made coincident.

If we introduce the following two conjugate circulant matrices,

$$\mathbf{V}=\begin{bmatrix} 1 & k^{-1} & \cdots\cdots & k \\ k & 1 & \cdots\cdots & k^2 \\ k^2 & k & \cdots\cdots & k^3 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdots\cdots & \cdot \\ \cdot & \cdot & & \cdot \\ k^{-1} & k^{-2} & \cdots\cdots & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{V}^*=\begin{bmatrix} 1 & k & \cdots\cdots & k^{-1} \\ k^{-1} & 1 & \cdots\cdots & k^{-2} \\ k^{-2} & k^{-1} & \cdots\cdots & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdots\cdots & \cdot \\ \cdot & \cdot & & \cdot \\ k & k^2 & \cdots\cdots & 1 \end{bmatrix}$$

wherein

$$k=\cos\theta_0+j\sin\theta_0,\ k^2=\cos 2\theta_0+j\sin 2\theta,\ k^{n-1}=k^{-1}=\cos\theta_0-j\sin\theta_0,$$

then equation (**5a**) may be rewritten as

$$\mathbf{Z}_{aa0}=\{R_a+p(L_{aa}-M_a')\}\mathbf{I}+p\frac{M_a'}{2}\{\mathbf{V}+\mathbf{V}^*\}, \quad \cdots\cdots \ (\mathbf{5b})$$

I being the identity matrix.

As for the saliency matrix, $\mathbf{Z}_{aa2}$, the maximum value of the second harmonic of the self-inductance of armature or stator per phase is taken to be equal to the maximum value of the second harmonic of the mutual inductance between armature phases (Kirschbaum 1950), so that essentially sinusoidal voltages under steady-state balanced operation are obtained. Sah (1939) has shown that for three-phase machines, the departure from equality gives an indication of the presence of third harmonics.

$\mathbf{Z}_{aa2}$ in non-ion form is

$$\mathbf{Z}_{aa2}=pL_{a2}'\begin{bmatrix} \cos 2\theta & \cos(2\theta-\theta_0) & \cdots\cdots\cdots & \cos(2\theta+\theta_0) \\ \cos(2\theta-\theta_0) & \cos(2\theta-2\theta_0) & \cdots\cdots\cdots & \cos 2\theta \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \cos(2\theta+\theta_0) & \cos 2\theta & \cdots\cdots\cdots & \cos(2\theta+\theta 2_0) \end{bmatrix}, \quad \cdots \ (\mathbf{6a})$$

where L_{a2}' is the maximum value of the second harmonic of the impedance, self or mutual, of armature per phase.

The angle, θ, is equal to ωt, where ω represents the rotational angular velocity between the armature and the field windings. The value of ω will be equal to $\omega_1=2\pi f$ for synchronous operation, but it may differ from ω_1 by an amount, S', for synchronous-induction operation.

In terms of the following two conjugate matrices:

$$\mathbf{W}=\begin{bmatrix} 1 & k & \cdots\cdots & k \\ k^{-1} & k^{-2} & \cdots\cdots & 1 \\ k^{-2} & \vdots & & \vdots \\ \vdots & \vdots & \cdots\cdots & \vdots \\ \vdots & \vdots & & \vdots \\ k & 1 & \cdots\cdots & k^{2} \end{bmatrix} \quad \text{and} \quad \mathbf{W}^{*}=\begin{bmatrix} 1 & k & \cdots\cdots & k^{-1} \\ k & k^{2} & \cdots\cdots & 1 \\ k^{2} & \vdots & & \vdots \\ \vdots & \vdots & \cdots\cdots & \vdots \\ \vdots & \vdots & & \vdots \\ k^{-1} & 1 & \cdots\cdots & k^{-2} \end{bmatrix}$$

$\mathbf{Z}_{aa2}$ may be expressed simply as

$$\mathbf{Z}_{aa2}=\frac{p}{2}L_{a2}'(\mathbf{W}\varepsilon^{j2\theta}+\mathbf{W}^{*}\varepsilon^{-j2\theta}). \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad \textbf{(6b)}$$

IV. ARMATURE-REACTION MATRIX

According to the two-reaction theory (Ku 1937; Park 1929), the mutual-inductance matrix between the stator and the rotor may be expressed by

$$\mathbf{M}_{fa}=\begin{bmatrix} M_{d1}\cos\theta & M_{d1}\cos(\theta-\theta_0) & \cdots\cdots & M_{d1}\cos(\theta+\theta_0) \\ M_{d2}\cos\theta & M_{d2}\cos(\theta-\theta_0) & \cdots\cdots & M_{d2}\cos(\theta+\theta_0) \\ \vdots & \vdots & \vdots & \vdots \\ M_{dm}\cos\theta & M_{dm}\cos(\theta-\theta_0) & \cdots\cdots & M_{dm}\cos(\theta+\theta_0) \\ -M_{q1}\cos\theta & -M_{q1}\cos(\theta-\theta_0) & \cdots\cdots & -M_{q1}\cos(\theta+\theta_0) \\ -M_{q2}\cos\theta & -M_{q2}\cos(\theta-\theta_0) & \cdots\cdots & -M_{q2}\cos(\theta+\theta_0) \\ \vdots & \vdots & \vdots & \vdots \\ -M_{qm}'\cos\theta & -M_{qm}'\cos(\theta-\theta_0) & \cdots\cdots & -M_{qm}'\cos(\theta+\theta_0) \end{bmatrix}, \ldots \textbf{(7a)}$$

and its transpose is $\mathbf{M}_{af}$. M_d is the maximum value of the mutual inductance between one phase of the armature and the direct-axis field winding, and M_q that between one phase of the armature and the quadrature-axis field winding. There are m circuits in the direct axis and m' circuits in the quadrature axis. Evidently $\mathbf{M}_{fa}$ may also be written as

$$\mathbf{M}_{fa}=\frac{\varepsilon^{j\theta}}{2}\begin{bmatrix} M_{d1} & M_{d1}k^{-1} & \cdots & M_{d1}k \\ \vdots & \vdots & \vdots & \vdots \\ M_{dm} & M_{dm}k^{-1} & \cdots & M_{dm}k \\ jM_{q1} & jM_{q1}k^{-1} & \cdots & jM_{q1}k \\ \vdots & \vdots & \vdots & \vdots \\ jM_{qm'} & jM_{qm'}k^{-1} & \cdots & jM_{qm'}k \end{bmatrix}+\frac{\varepsilon^{-j\theta}}{2}\begin{bmatrix} M_{d1} & M_{d1}k & \cdots & M_{d1}k^{-1} \\ \vdots & \vdots & \vdots & \vdots \\ M_{dm} & M_{dm}k & \cdots & M_{dm}k^{-1} \\ -jM_{q1} & -jM_{q1}k & \cdots & -jM_{q1}k^{-1} \\ \vdots & \vdots & \vdots & \vdots \\ -jM_{qm'} & -jM_{qm'}k & \cdots & -jM_{qm'}k^{-1} \end{bmatrix}.$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \textbf{(7b)}$$

The self-impedance matrix for the field system is

$$\mathbf{Z}_{ff}=\left[\begin{array}{cccccccc} R_{df1}+pL_{df1} & pM_{df12} & \cdots & pM_{df1m} & & & & \\ pM_{df12} & R_{df2}+pL_{df2} & \cdots & \vdots & & & & \\ \vdots & \vdots & \vdots & \vdots & & & & \\ pM_{df1m} & \cdots & \cdots & R_{dfm}+pL_{dfm} & & & & \\ & & & & R_{qf1}+pL_{qf1} & pM_{qf12} & \cdots & pM_{qf1m'} \\ & & & & pM_{qf12} & R_{qf2}+pL_{qf2} & \cdots & \cdots \\ & & & & \vdots & \vdots & & \vdots \\ & & & & pM_{qf1m'} & \cdots & \cdots & R_{qfm'}+pL_{qfm'} \end{array}\right], \quad \ldots\ldots \quad (8a)$$

wherein R_{df} and L_{df} are the resistance and self-inductance of direct-axis field windings, and M_{df} the mutual inductance between direct-axis field windings. R_{qf}, L_{qf}, and M_{qf} are the corresponding constants of the quadrature-axis field windings.

Writing $\mathbf{Z}_{ffd}$ for the $m\,x\,m$ sub-matrix at the upper left of (**8a**), and $\mathbf{Z}_{ffq}$ for the $m'\,x\,m'$ sub-matrix at the lower right of (**8a**), and using partitioned matrices, we have

$$\mathbf{Z}_{ff}=\begin{bmatrix} \mathbf{Z}_{ffd} & 0 \\ 0 & \mathbf{Z}_{ffq} \end{bmatrix}. \quad \ldots\ldots\ldots\ldots \quad (8b)$$

Substituting equation (**4b**) into (**4a**), we have

$$\mathbf{V}_a=\mathbf{Z}_{aa}\mathbf{I}_a-p\mathbf{M}_{af}\mathbf{Z}_{ff}{}^{-1}p\mathbf{M}_{fa}\mathbf{I}_a+pM_{af}\mathbf{Z}_{ff}{}^{-1}\mathbf{E}_f, \quad \ldots\ldots \quad (9a)$$

or

$$\mathbf{e}_a=\mathbf{v}_a-p\mathbf{M}_{af}\mathbf{Y}_{ff}\mathbf{E}_f=(\mathbf{Z}_{aa}-p\mathbf{M}_{af}\mathbf{Y}_{ff}p\mathbf{M}_{fa})\mathbf{I}_a, \quad \ldots\ldots \quad (9b)$$

where

$$\mathbf{Y}_{ff}=\mathbf{Z}_{ff}{}^{-1}$$

$$=\begin{bmatrix} \mathbf{Z}_{ffa}{}^{-1} & 0 \\ 0 & \mathbf{Z}_{ffq}{}^{-1} \end{bmatrix}. \quad \ldots\ldots\ldots\ldots \quad (10)$$

The quantity within the parentheses on the right-hand side of equation (**9b**) will be called the operational impedance matrix of the machine and the term, $p\mathbf{M}_{af}\mathbf{Y}_{ff}p\mathbf{M}_{fa}$, gives the reaction of the rotor on the stator. It should be noted that the operator p standing out explicitly or implicitly in $\mathbf{Y}_{ff}$ applies to every quantity written to its right. The above differential equations with variable coefficients will be simplified by the following transformations.

V. AN AFFINE TRANSFORMATION

A careful study of the forms of the self- and mutual-impedance matrices suggests the two following unitary matrices to be used as an affine transformation of space with origin fixed.

$$\mathbf{U}=\frac{1}{\sqrt{n}}\begin{bmatrix} 1 & 1 & 1 & \ldots\ldots & 1 \\ \varepsilon^{-j\theta} & k\varepsilon^{-j\theta} & k^2\varepsilon^{-j\theta} & \ldots\ldots & k^{-1}\varepsilon^{-j\theta} \\ 1 & k^2 & k^4 & \ldots\ldots & k^{-2} \\ 1 & k^3 & k^6 & \ldots\ldots & k^{-3} \\ \cdot & \cdot & \cdot & & \vdots \\ \varepsilon^{j\theta} & k^{-1}\varepsilon^{j\theta} & k^{-2}\varepsilon^{j\theta} & \ldots\ldots & k\varepsilon^{j\theta} \end{bmatrix}, \quad \ldots \quad \mathbf{(11a)}$$

$$\mathbf{U}^{-1}=\mathbf{U}_t^{*}=\frac{1}{\sqrt{n}}\begin{bmatrix} 1 & \varepsilon^{j\theta} & 1 & \ldots\ldots & \varepsilon^{-j\theta} \\ 1 & k^{-1}\varepsilon^{j\theta} & k^{-2} & \ldots\ldots & k\varepsilon^{-j\theta} \\ 1 & k^{-2}\varepsilon^{j\theta} & k^{-4} & \ldots\ldots & k^2\varepsilon^{-j\theta} \\ \cdot & \cdot & \cdot & & \vdots \\ 1 & k\varepsilon^{j\theta} & k^2 & \ldots\ldots & k^{-1}\varepsilon^{-j\theta} \end{bmatrix}. \quad \ldots \quad \mathbf{(11b)}$$

It should be noted that the new coordinates are also quasi-coordinates which uniquely determine the system and the transformation is always reversible. Thus, the new voltage and current vectors are

$$\mathbf{V}'=\mathbf{U}\mathbf{V}_a=\{v_u\ v_s\ v_v\ v_w\ldots .v_n\}, \quad \ldots\ldots\ldots\ldots \quad \mathbf{(12)}$$

and

$$\mathbf{I}'=\mathbf{U}\mathbf{I}_a=\{i_u\, i_s\, i_v\, i_w\ldots .i_n\}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \mathbf{(13)}$$

Here, u, v, w, ..., etc., are stationary axes, whereas s and n are rotating axes or variable base vectors.

Since **U** is non-singular, the constant part of the self-impedance matrix (equation (**5b**)) is diagonalized by the following operation :

$$\mathbf{U}\mathbf{Z}_{aa0}\,\mathbf{U}_t^{*}=\begin{bmatrix} R_a+p(L_{aa}-M_a') & 0 & 0 & \ldots . 0 \\ 0 & R_a+(p+j\omega)(L_{aa}+\frac{n-2}{2}m_a') & 0 & \ldots . 0 \\ 0 & 0 & R_a+p(L_{aa}-m_a') & \ldots . 0 \\ \cdot & \cdot & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & R_a+(p-j\omega)\left(L_{aa}+\frac{n-2}{2}m_a'\right) \end{bmatrix}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \mathbf{(14)}$$

in which all elements in the principal diagonal except the second and the last are $R_a+p(L_{aa}-m_a')$. Now if we write $L_a=L_{aa}-m_a'$ and $L_{a0}=L_{aa}+\left(\frac{n}{2}-1\right)m_a'$, L_a will be the constant part of the self-inductance of the armature phase winding, and L_{a0} the constant self-inductance part of one phase plus the constant mutual effect of the other phases ; therefore, L_{a0} may be defined as the constant part of the synchronous self-inductance of the armature per phase.

By the same transformation, the saliency impedance matrix, $\mathbf{Z}_{aa2}$, as given in equation (**6b**) becomes

$$\mathbf{U}\mathbf{Z}_{aa2}\mathbf{U}_t^* = \begin{bmatrix} 0 & 0 & 0 \;\ldots\ldots\ldots & 0 \\ 0 & 0 & \cdot & (p+j\omega)L_{a2} \\ \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & (p-j\omega)L_{a2} & 0 \;\ldots\ldots\ldots & 0 \end{bmatrix} . \quad .. \; (\mathbf{15})$$

It is interesting to note that the saliency effect depends only upon the two rotating axes given by the second and last rows of the matrix **U**. Physically, due to the saliency of poles, a positively-rotating voltage vector does not give rise to a positively-rotating current vector only, and conversely a positively-rotating current will, in general, depend upon both the positively- and negatively-rotating voltages.

To carry out the same transformation to the armature reaction matrix, $p\mathbf{M}_{af}\mathbf{Y}_{ff}p\mathbf{M}_{fa}$, we notice that $\mathbf{M}_{af}$ and $\mathbf{M}_{fa}$ are transpose of each other, pre-multiplying $p\mathbf{M}_{af}$ by **U** giving only two rows, the second and the last, different from zero, and post-multiplying $p\mathbf{M}_{fa}$, by $\mathbf{U}_t^*$ gives only two columns, the second and the last, different from zero. $\mathbf{Y}_{ff}$ does not affect the form of the resultant matrix, though its elements are complicated algebraic rational functions of p.

In terms of the sub-matrices, $\mathbf{Z}_{ffd}$ and $\mathbf{Z}_{ffq}$, and the following row matrices :

$$\mathbf{M}_{d_t} = (M_{d1}, \ldots . M_{dm}), \qquad \mathbf{M}_{q_t} = (M_{q1}, \ldots . M_{qm}{}'),$$

and their transposes, $\mathbf{M}_d$ and $\mathbf{M}_q$, the transformation results in the following single form :

$$\mathbf{U}\,p\,\mathbf{M}_{af}\mathbf{Y}_{ff}\,p\,\mathbf{M}_{fa}\mathbf{U}_t^* = \begin{bmatrix} 0 & 0 & 0 \;\ldots\ldots\ldots & 0 \\ 0 & T(p) & 0 \;\ldots\ldots\ldots & R(p) \\ 0 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & R^*(p) & 0 \;\ldots\ldots\ldots & T^*(p) \end{bmatrix}, \quad .. \; (\mathbf{16})$$

where $$T(p)=\frac{n}{2}(p+j\omega)(\mathbf{M}_{d_t}\mathbf{Z}_{ff}^{-1}\,p\,\mathbf{M}_d+\mathbf{M}_{qt}\,\mathbf{Z}_{ff}^{-1}\,p\,\mathbf{M}_q), \quad \ldots\ldots \quad \textbf{(17a)}\dagger$$

$$R(p)=\frac{n}{2}(p+j\omega)(\mathbf{M}_{d_t}\,\mathbf{Z}_{ff}^{-1}\,p\,\mathbf{M}_d-\mathbf{M}_{qt}\,\mathbf{Z}_{ff}^{-1}\,p\,\mathbf{M}_q). \quad \ldots\ldots \quad \textbf{(17b)}$$

$T^*(p)$ and $R^*(p)$ are the corresponding conjugates.

The above operational functions can be easily determined for particular cases.

As for the matrix due to D.C. potentials in the field circuit, that is $p\,\mathbf{M}_{af}\mathbf{Y}_{ff}\mathbf{E}_f$, pre-multiplication by $\mathbf{U}$ yields a column matrix with only two conjugate elements :

$$\mathbf{U}\,p\,\mathbf{M}_{af}\mathbf{Y}_{ff}\mathbf{E}_f=\{0,\,H(p)\;0\;\;0\;\ldots.\;H^*(p)\}, \quad \ldots\ldots\ldots \quad \textbf{(18)}$$

in which

$$H(p)=\frac{\sqrt{n}}{2}(p+j\omega)(\mathbf{M}_{d_t}\mathbf{Z}_{ffd}^{-1}\mathbf{E}_{df}+j\,\mathbf{M}_{q_t}\mathbf{Z}_{ffq}^{-1}\mathbf{E}_{qf}). \quad \ldots\ldots \quad \textbf{(19)}$$

is expressed in terms of the column matrices

$$\mathbf{E}_{df}=\{E_{df1},\ldots.E_{dfm}\} \text{ and } \mathbf{E}_{qf}=\{E_{qf1},\ldots.E_{qfm}'\}.$$

$\mathbf{M}_{d_t}\mathbf{Z}_{ffd}^{-1}$ and $\mathbf{M}_{q_t}\mathbf{Z}_{ffd}^{-1}$ are the operational direct and quadrature conductance matrices (Park 1929).

By defining the two following operational inductances :

$$C(p)=L_{a0}-\frac{T(p)}{(p+j\omega)}, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad \textbf{(20a)}$$

$$S(p)=L_{a2}-\frac{R(p)}{(p+j\omega)}, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad \textbf{(20b)}$$

equation (**9b**) can now be expressed in the following two groups after the unitary transformation :

$$\{v_u,\,v_v,\,v_w,\ldots.\}=(R_a+pL_a)\{i_u,\,i_v,\,i_w,\ldots.\}, \quad \ldots\ldots \quad \textbf{(21a)}$$

and

$$\begin{bmatrix} e_s \\ e_n \end{bmatrix}=\begin{bmatrix} v_s \\ v_m \end{bmatrix}-\begin{bmatrix} H(p) \\ H^*(p) \end{bmatrix}=\begin{bmatrix} R_a+(p+j\omega)C(p) & (p+j\omega)S(p) \\ (p-j\omega)S(p) & R_a+(p-j\omega)C(p) \end{bmatrix}\begin{bmatrix} i_s \\ i_n \end{bmatrix}. \quad \ldots \quad \textbf{(21b)}$$

It is interesting to note that the currents, i_u, i_v, i_w, ..., etc., form a group and do not involve the field quantities. The author proposes to call them " zero armature reaction currents ". In geometry, these $n-2$ axes form a sub-space orthogonal to the two-dimensional plane on which the two rotating axes s and n lie. In the special case of three-phase machines, the $n-2$ sub-space becomes simply an isoclinic vector describing the zero-sequence system, hence it is possible to define L_a also as a zero-sequence inductance of the machine. The two rotating axes then lie on a plane which is perpendicular to this isoclinic vector passing through the origin. The locus described on this plane will be an ellipse, but may degenerate to a circle if the system is balanced.

---

† The author is indebted to Professor E. R. Love, Department of Mathematics, University of Melbourne, for assistance in putting equations (**17**) and (**19**) in the present form.

VI. Transient Solutions

The transient solutions for all the zero armature reaction components can be easily obtained from (**21a**) and then interpreted by means of Laplacian transform. For suddenly applied voltages, it will be assumed that all field windings are initially short-circuited, i.e. $\mathbf{E}_{df}$ and $\mathbf{E}_{qf}$ are equal to zero. Then from equation (**21b**), we have

$$\begin{bmatrix} i_s \\ i_n \end{bmatrix} = \frac{1}{A(p)} \begin{bmatrix} R_a+(p-j\omega)C(p) & -(p+j\omega)S(p) \\ -(p-j\omega)S(p) & R_a+(p+j\omega)C(p) \end{bmatrix} \begin{bmatrix} V_s \\ V_n \end{bmatrix}, \quad \ldots \quad (\mathbf{22})$$

where $$A(p)=R_a{}^2+2R_a\,p\,C(p)+(p^2+\omega^2)\{[C(p)]^2-[S(p)]^2\}. \quad \ldots\ldots \quad (\mathbf{23})$$

It may be noted that if V_s and V_n are conjugate expressions, as is usually the case, i_s and i_n will be conjugate and hence only one variable need be evaluated.

If we introduce the transformation matrix $\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{j}{\sqrt{2}} & \frac{j}{\sqrt{2}} \end{bmatrix}$

such that

$$\begin{bmatrix} V_d \\ V_q \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{j}{\sqrt{2}} & \frac{j}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} V_s \\ V_n \end{bmatrix}, \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{24})$$

equation (**22**) becomes

$$\begin{bmatrix} i_d \\ i_q \end{bmatrix} = \frac{1}{A(p)} \begin{bmatrix} R_a+pQ(p) & \omega Q(p) \\ -\omega D(p) & R_a+pD(p) \end{bmatrix} \begin{bmatrix} V_d \\ V_q \end{bmatrix}, \quad \ldots\ldots \quad (\mathbf{25})$$

where

$$\left.\begin{aligned} D(p)&=C(p)+S(p), \\ Q(p)&=C(p)-S(p). \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{26})$$

In terms of $D(p)$ and $Q(p)$, $A(p)$ becomes

$$A(p)=R_a{}^2+R_a\,p[D(p)+Q(p)]+(p^2+\omega^2)[D(p)][Q(p)]. \quad \ldots \quad (\mathbf{27})$$

For steady-state synchronous operation,

$$p=0,\ C(p)=L_a,\ S(p)=L_{a2},\ A(p)=R_a{}^2+\omega^2(L_{a0}{}^2)-L_{a2}{}^2),$$

we may define

$L_{ad}=D(0)=L_{a0}+L_{a2}$ as the direct-axis synchronous inductance,

and

$L_{aq}=Q(0)=L_{a0}-L_{a2}$ as the quadrature-axis synchronous inductance.

The correlation of our solution with that by double-reaction theory is thus established. Physically, the double-reaction theory transforms the stationary armature winding into two independent windings in space quadrature both rotating synchronously with the rotor.

To deal with sudden short-circuit, a set of armature voltages equal to the negative value of the original set of voltages are suddenly applied to the armature terminals, and these new currents are then superimposed on the original steady-state distribution of currents to get the final results. In this case V_s and V_n in equation (**22**) must be opposite in sign to the original voltages before short-circuit.

The original steady-state current due to $-e_s$ and $-e_n$ is

$$\begin{bmatrix} -i_s \\ \\ -i_n \end{bmatrix}_{s.s.} = \frac{1}{R_a{}^2+X_{0a}{}^2-X_{2a}{}^2} \begin{bmatrix} R_a-jX_{0a} & -jX_{a2} \\ \\ -jX_{a2} & R_a+jX_{0a} \end{bmatrix} \begin{bmatrix} e_s \\ \\ e_n \end{bmatrix}, \quad \text{. . (28a)}$$

or

$$\begin{bmatrix} -i_d \\ \\ -i_q \end{bmatrix}_{s.s.} = \frac{1}{R_a{}^2+X_{ad}X_{aq}} \begin{bmatrix} R_a & X_{aq} \\ \\ -X_{ad} & R_a \end{bmatrix} \begin{bmatrix} e_d \\ \\ e_q \end{bmatrix}. \quad \text{. . . . (28b)}$$

In the case of symmetrical short-circuit under no-load, V_s and V_n are not only conjugate, but also one is equal to the negative of the other, and equation (**22**) simplifies to

$$\begin{bmatrix} i_s \\ \\ i_n \end{bmatrix} = \begin{bmatrix} \dfrac{R_a+pD(p)-j\omega Q(p)}{A(p)} & 0 \\ 0 & \dfrac{R_a+pD(p)+j\omega Q(p)}{A(p)} \end{bmatrix} \begin{bmatrix} V_s \\ \\ V_n \end{bmatrix}. \quad \text{. . (29)}$$

For short-circuit under load, the induced e.m.f. will differ from the terminal voltage by an angle, α, say. In the case of the voltages being balanced, V_s and V_n will be of the form $\varepsilon^{-j\alpha}$ and $\varepsilon^{j\alpha}$ respectively. As these voltage components are conjugate, the current components will also be conjugate. If the resistance is neglected, equation (**25**) can be simplified to

$$\begin{bmatrix} i_d \\ \\ i_q \end{bmatrix} = \frac{1}{(p^2+\omega^2)D(p)Q(p)} \begin{bmatrix} pQ(p) & Q(p) \\ -D(p) & pD(p) \end{bmatrix} \begin{bmatrix} V_d \\ \\ V_q \end{bmatrix}. \quad \text{. . (30)}$$

VII. OPERATIONAL FUNCTIONS FOR SIMPLIFIED FIELD STRUCTURE

Let the field structure and windings of the n-phase machine be simplified so that they will be the same as a general three-phase machine, viz. the main winding, field collar, and damper winding. The action of short-circuited collar will be taken as equivalent to a winding having the same axis as the main field (Waring and Crary 1932). A damper winding is taken as being equivalent to two windings, one having the same axis as the main field and the other in space quadrature with it. Let the main winding, field collar, and equivalent direct-axis damper be numbered 1, 2, 3, respectively. The self-impedance matrix of the field system, $\mathbf{Z}_{ff}$, in equation (**8a**) becomes

$$\mathbf{Z}_{ff} = \begin{bmatrix} R_{df1}+p\,L_{df1} & pM_{df12} & pM_{df13} & 0 \\ pM_{df12} & R_{df2}+p\,L_{df2} & pM_{df23} & 0 \\ pM_{df13} & pM_{df23} & R_{df3}+p\,L_{df3} & 0 \\ 0 & 0 & 0 & R_{qf1}+p\,L_{qf1} \end{bmatrix}, \quad \text{. . (8c)}$$

and its inverse is

$$\mathbf{Y}_{ff}=\begin{bmatrix} Y_{11} & Y_{12} & Y_{13} & 0 \\ Y_{12} & Y_{22} & Y_{23} & 0 \\ Y_{13} & Y_{23} & Y_{33} & 0 \\ 0 & 0 & 0 & (R_{qf1}+p\,L_{qf1})^{-1} \end{bmatrix}, \quad \dots\dots\dots\dots \quad \mathbf{(8d)}$$

wherein

$$Y_{11}=\frac{1}{A_d(p)}[(R_{df2}+p\,L_{df2})(R_{df3}+p\,L_{df3})-M_{df}{}^2{}_{23}]$$

$$Y_{22}=\frac{1}{A_d(p)}[(R_{df1}+p\,L_{df1})(R_{df3}+p\,L_{df3})-M_{df}{}^2{}_{13}]$$

$$Y_{33}=\frac{1}{A_d(p)}[R_{df1}+p\,L_{df1})(R_{df2}+p\,L_{df2})-M_{df}{}^2{}_{12}]$$

$$Y_{12}=\frac{1}{A_d(p)}[M_{df13}M_{df23}-(R_{df3}+p\,L_{df3})M_{df12}]$$

$$Y_{13}=\frac{1}{A_d(p)}[M_{df12}M_{df23}-(R_{df2}+p\,L_{df2})M_{df13}]$$

$$Y_{23}=\frac{1}{A_d(p)}[M_{df12}M_{df13}-(R_{df1}+p\,L_{df1})M_{df23}]$$

and $A_d(p)$ is the determinant of the third order sub-matrix in equation (**8d**).

Using $\mathbf{Y}_{ff}$ of (**8c**) in the operation, $\mathbf{U}\,p\,M_{af}\mathbf{Y}_{ff}p\,\mathbf{M}_{f_a}\mathbf{U}_t{}^*$, we obtain

$$\frac{n}{2}\mathbf{M}_{d_t}\mathbf{Z}_{ff}{}^{-1}\mathbf{M}_d=\frac{n}{2}\,p\,[M_{d_1}\alpha_1(p)+M_{d_2}\alpha_2(p+M_{d_3}\alpha_3(p)], \quad \dots\dots\dots\dots \quad \mathbf{(31)}$$

where

$$\alpha_1(p)=M_{d_1}Y_{11}+M_{d_2}Y_{12}+M_{d_3}Y_{13}$$
$$\alpha_2(p)=M_{d_1}Y_{12}+M_{d_2}Y_{22}+M_{d_3}Y_{23}$$
$$\alpha_3(p)=M_{d_1}Y_{13}+M_{d_2}Y_{23}+M_{d_3}Y_{33}.$$

As there is only one rotor winding in the quadrature field axis as previously assumed, hence

$$\frac{n}{2}\mathbf{M}_{q_t}\mathbf{Z}_{ff}{}^{-1}\mathbf{M}_q=\frac{n}{2}\,p\left[\frac{M_{q1}{}^2}{R_{qf1}+p\,L_{qf1}}\right].$$

Similarly, by performing $\mathbf{U}p\mathbf{M}_{af}\mathbf{Y}_{ff}\mathbf{E}_f$, we find that the elements of $\sqrt{n}\mathbf{M}_{d_t}\mathbf{Z}_{ffd}{}^{-1}$ are

$$2(M_{d_1}Y_{1k}+M_{d_2}Y_{2k}+M_{d_3}Y_{3k}), \qquad k=1,\ 2,\ 3.$$

It is not necessary to determine $\sqrt{n}\mathbf{M}_{q_t}\mathbf{Z}_{ffd}{}^{-1}$ if the quadrature winding is short-circuited.

When p approaches infinity, or for the resistance to be zero, $L_{ad}(p)$ will reduce to $L_d{}''$, the so-called sub-transient inductance of the direct-axis armature circuit. We have then

$$L_d{}''=\underset{p\to\infty}{D(p)}=L_{ad}-M_d{}''. \quad \dots\dots\dots\dots \quad \mathbf{(32)}$$

The expression for $M_d{}''$ is rather lengthy for machines with rotor windings consisting of amortisseur, short-circuited collar, and the main field winding, but can be obtained easily from equation (**31**).

If there is only one additional rotor winding besides the main field winding in the direct axis, we have

$$L_d''=\underset{p\to\infty}{D(p)}=L_{ad}-\frac{n}{2}\left[\frac{M_{d1}{}^2L_{df2}+M^2{}_{d2}L_{df1}-2M_{d1}M_{d2}M_{df12}}{L_{df1}L_{df2}-M_{df}{}^2{}_{12}}\right]. \quad \ldots\ldots \quad (33)$$

The above differs from that given by Waring and Crary (1932) by a factor $n/2$, since the constants in their paper were in per unit values in terms of the armature circuit.

If there is no additional winding in the direct axis, we have simply

$$\frac{n}{2}\mathbf{M}_{dt}\mathbf{Z}_{ff}{}^{-1}\mathbf{M}_d=\frac{n}{2}\,p\left[\frac{M_{d1}{}^2}{R_{df1}+p\,L_{df1}}\right].$$

We get the transient inductance (Deborne 1950) value as

$$L_d'=\underset{p\to\infty}{D(p)}=L_{ad}-\frac{n}{2}\left[\frac{M_{d1}{}^2}{L_{df1}}\right]=\sigma_dL_{ad}, \quad \ldots\ldots\ldots\ldots \quad (34)$$

where

$$\sigma_d=1-\frac{n}{2}\left[\frac{M_{d1}{}^2}{L_{ad}L_{df1}}\right].$$

The transient inductance in the quadrature field axis will be

$$L_q'=\underset{p\to\infty}{Q(p)}=L_{aq}-\frac{n}{2}\,\frac{M_{q1}{}^2}{L_{qf}}=\sigma_qL_{aq}, \quad \ldots\ldots\ldots\ldots \quad (35)$$

where

$$\sigma_q=1-\frac{n}{2}\,\frac{M_{q1}{}^2}{L_{aq}L_{qf1}}.$$

If only one winding is present in the quadrature field axis, L_q' will be equal to the sub-transient, L_q''.

VIII. CONCLUSIONS

The above analysis establishes the principle of "zero armature reaction currents" but fails in the case of single-phase machines because they are completely unsymmetrical. The least number of phases that the above analysis holds is three inasmuch as the so-called two-phase machine is strictly speaking a semi-four-phase one. The term, two-phase circuit, in which the voltages make 180° with each other, should properly be called a three-wire single-phase circuit (Shen 1950). It appears that, for ordinary physical problems and particularly practical ones, that the engineering profession must encounter, it is advisable to group the variables in such a way that the unimportant ones can be put aside for a while and that only three outstanding ones remain.

For completely-symmetrical machines, since windings on both sides of the air gap are polyphase and symmetrical, the two rotating axes, s and n, will be independent. In other words, a positively-rotating voltage vector will give rise to a positively-rotating current vector, and similarly for the negatively-rotating vectors. Finally for steady-state balanced operation, i_n will be the image current of i_s; thus equation (**21b**) reveals the interesting fact that the saliency of poles is responsible for an apparent armature phase shift, because the drop due to synchronous inductance involves armature current, while the saliency feature is accounted for by a voltage drop involving the second-harmonic inductance and the image armature current ; this fact is obscured if the analysis is based entirely on the two-reaction theory.

IX. ACKNOWLEDGMENTS

The author is indebted to Dr. W. J. Gibbs, of British Thomson Houston Co. Ltd., Rugby, and Professor E. O. Willoughby and Mr. R. E. Vowels, both of the University of Adelaide, for their interest and helpful criticism. He also wishes to thank the referees for their comments and suggestions.

X. REFERENCES

DEBORNE, A. (1950).—Inductances transitoires des machines synchrones à trois et A.N. enroulements. *Bull. Soc. Franc. Elect.* **10** : 223.

GIBBS, W. J. (1950).—Limitations of Lagrangian methods in electrical machine theory. *Beama J.* **57** : 342-9.

KIRSCHBAUM, H. S. (1950).—Per unit inductances of synchronous machines. *Trans. Amer. Inst. Elect. Engrs.* **69** : 231-4.

KRON, G. (1934).—Non-Riemannian dynamics of rotating electrical machinery. *J. Math. Phys.* **13** : 103.

KRON, G. (1935).—The application of tensors to the analysis of rotating electrical machinery. *Gen. Elect. Rev.* **38** : 39, 40.

KRON, G. (1936).—Quasi-holonomic dynamical systems. *Physics* **7** : 143.

KRON, G. (1942).—" Tensor Analysis." (John Wiley & Sons : New York.)

KRON, G. (1950).—Classification of the reference frames of a synchronous machine. *Trans. Amer. Inst. Elect. Engrs.* **69** : 720-7.

KU, Y. H. (1937).—Extension of two-reaction theory to multiphase synchronous machines. *Elect. Engr.* **56** : 1197.

LAUDNER, A. H. (1936).—Salient pole motors out of synchronism. *Elect. Engr.* **55** : 636.

PARK, R. H. (1928).—Definition of an ideal synchronous machine. *Gen. Elect. Rev.* **31** : 332.

PARK, R. H. (1929).—Two-reaction theory of synchronous machines. *Trans. Amer. Inst. Elect. Engrs.* **48** : 716.

SAH, A. P. T. (1936).—Dyadic algebra applied to three-phase circuits. *Elect. Engr.* **55** : 872.

SAH, A. P. T. (1939).—" Dyadic Circuit Analysis." (International Textbook Co. : Scranton, Penn.)

SHEN, D. W. C. (1948).—Generalized co-ordinates in substitution networks. *Phil. Mag.* **39** : 890.

SHEN, D. W. C. (1950).—A.C. circuits. Electrical Engineering Department, University of Adelaide (mimeo.).

VOWELS, R. E. (1944).—Transformation matrices in electrical engineering. *Trans. Instn. Engrs. Aust.* **25** : 97.

WARING, M. L., and CRARY, S. B. (1932).—The operational impedances of a synchronous machine. *Gen. Elect. Rev.* **35** : 578.

AN ANALYSIS OF THE GALVANOMETER AMPLIFIER AND ITS RESPONSE TO ALTERNATING ELECTROMOTIVE FORCES AND MECHANICAL VIBRATIONS

By R. G. Wylie* and A. F. A. Harper*

[*Manuscript received June 29, 1951*]

Summary

A system which is representative of most practical galvanometer amplifiers is analysed and the response of the system to alternating e.m.f.'s and to mechanical vibration is determined.

It is shown that the values of the circuit constants can be so chosen that the galvanometer movement is at least critically damped for all conditions of operation and that, for critical damping, the frequency response curve for the amplifier to applied alternating e.m.f.'s resembles that for the galvanometer alone but is expanded along the frequency axis by a factor which depends on the degree of feedback. Mechanical vibration with a component about the axis of the galvanometer coil produces an effect which is markedly increased by the provision of feedback.

I. Introduction

Of several devices developed in recent years for the amplification of very small D.C. e.m.f.'s the galvanometer amplifier employing negative feedback has been one of the most successful. This type of amplifier, which was described by Gall in 1942, incorporates a reflecting galvanometer which is optically linked to either one or two phototubes and, when suitably designed and operated under good conditions, has a zero stability of about $\pm 0 \cdot 01\ \mu V$. It possesses the desirable characteristics common to all feedback amplifiers together with some unusual properties associated with the kinetics of the galvanometer.

Practical forms of the amplifier have been described by Gall (1942), Preston (1946), Hill (1948), Frankenhaeuser and MacDonald (1949), and in a report by one of the present authors (1951). In that report a description of an amplifier employing two phototubes is given with an account of its measured performance; particular attention is given to zero drift.

The possibilities and limitations of the galvanometer amplifier may be studied by analysing a typical system. The significant characteristics which such an analysis should give are the amplification, the effective input and output resistances, the damping, the response to a varying input e.m.f., and the effect on the system of mechanical vibration. Special interest attaches to those properties, not common to all types of feedback amplifier, which arise from the presence of the galvanometer. Some of these properties have been considered in an analysis given by Gall.

---

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

The present paper discusses the general properties of the galvanometer amplifier in some detail, special attention being given to the damping properties when the galvanometer has a permanent shunting resistance, and the effect of vibration. These have not previously been adequately examined. It is shown that the values of the circuit constants can be so chosen that the galvanometer movement is at least critically damped for all conditions of operation and that, for critical damping, the frequency response curve for the amplifier to applied alternating e.m.f.'s resembles that for the galvanometer alone but is expanded along the frequency axis by a factor which depends on the feedback in the system. The effect of mechanical vibrations with a rotational component about the axis of the galvanometer coil is markedly increased by the provision of feedback.

II. The General Circuit

Figure 1 is a block diagram which is capable of representing most practical galvanometer amplifiers. In the figure, R_g and R_s are respectively the resistances of the galvanometer and the permanent galvanometer shunt, and R is the internal resistance of the source of e.m.f. R_1 and R_2 are two resistors of stable resistance and R_1 is the resistance of the load. The quantities V, V_1, V_2, V_f, and V_0 are the potential differences between the points indicated ; the convention of signs followed is that V, V_1, V_2, and V_0 are positive, and V_f is negative when the polarities from top to bottom are positive-negative.

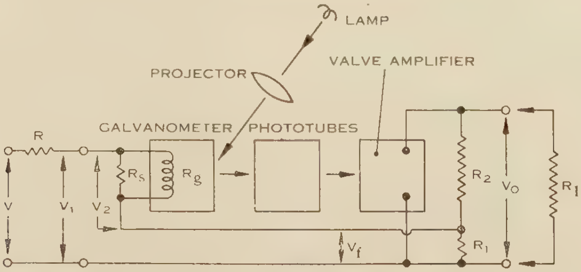


Fig. 1.—Block diagram of a galvanometer feedback amplifier.

The assumptions are made that the internal resistance presented by the valve amplifier to the resistance (R_1+R_2) is large compared with (R_1+R_2) and that both the self-inductance of the galvanometer coil and the time constant of the phototube circuit are not significant for the frequencies at which the instrument would normally be used.

III. Analysis for Stationary Galvanometer Coil

When a D.C. e.m.f., V, is applied to the amplifier a steady state associated with a steady output p.d., V_0, is established after an initial transient. For the analysis of this steady state and for subsequent sections it is convenient to define gain factors m, m_1, and m_2 by

$$m=\frac{V_0}{V},\ m_1=\frac{V_0}{V_1},\ \text{and}\ m_2=\frac{V_0}{V_2},$$

it being assumed that no zero error is present, i.e. that $V_0=0$, when $V=0$. Also, quantities β and β_0 may be defined by

$$\beta=\frac{V_f}{V_0} \text{ and } \beta_0=-\frac{R_1}{R_1+R_2}.$$

It will be clear that the gain factor m_2 is not influenced by the feedback p.d., V_f, whereas m_1 is influenced by V_f, and m is further influenced by the source resistance R. It is m_1 which would normally be called the gain of the amplifier. In practice β and β_0 are approximately equal, since R_1 is small compared with R_2, R_g, and R_s.

It is easily shown that the effective gain m_1 is given by

$$m_1=\frac{m_2}{1-\beta m_2}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

and that R_i, the effective internal resistance of the amplifier to an applied p.d., V_1, is given by the expression

$$R_i=(1\quad\beta m_2)X, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

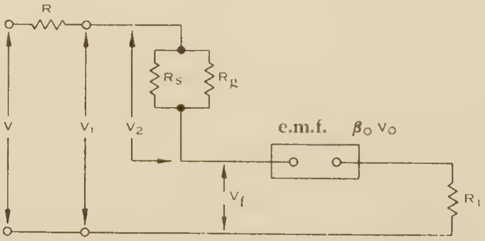


Fig. 2.—Circuit equivalent to the block diagram of Figure 1.

where X represents the resistance of R_s and R_g in parallel. These results, which are well known, are exact. An effective output resistance, R_0, may be defined by the equation

$$\frac{V_0'}{V_0}=\frac{R_l}{R_l+R_0},$$

where V_0' is the actual p.d. across the load resistance, R_l, and V_0 is the corresponding output p.d. when R_l is infinite. If the source resistance R is zero, R_0 is given approximately by

$$R_0=\frac{R_1+R_2}{(1-\beta m_2)}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

In practice both m_2 and $(1-\beta m_2)$ are large compared with unity, so that equation (1) can be written

$$m_1 \doteqdot -\frac{1}{\beta}\left(1+\frac{1}{\beta m_2}\right),$$

which shows that m_1 approximates to $-1/\beta$ which in turn approximates to $-1/\beta_0$. The gain m_1 may thus be practically independent of m_2.

Since the quantity β is not known *a priori*, it is desirable to express the results in terms of the resistance ratio, β_0. Rather than obtain the relationship between β and β_0 by purely algebraic means we may, for infinite load resistance, represent the block diagram of Figure 1 by the equivalent circuit of Figure 2. This representation is satisfactory if $R_1 \ll (X+R)$; if in addition $R_1 \ll R_2$ and $m \gg 1$ the representation is very accurate indeed.

Regarding R and R_1 as part of an amplifier to which the feedback e.m.f., $\beta_0 V_0$, is applied it is seen from Figure 2 that

$$m = \frac{m_2 X/(X+R+R_1)}{1-\beta_0 m_2 X/(X+R+R_1)}. \quad \ldots\ldots\ldots\ldots\ldots \quad (4)$$

Putting $R=0$ in (**4**) to obtain m_1 and comparing with equation (**1**) we obtain

$$\frac{\beta}{\beta_0} = 1 - \frac{R_1}{m_2 \beta_0 X}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

Substituting for β in equation (**2**),

$$R_i = (1-\beta_0 m_2)X + R_1. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (6)$$

The corresponding expression for the effective output resistance is

$$R_0 = \frac{R_1+R_2}{1-\beta_0 m_2 X/(X+R+R_1)}. \quad \ldots\ldots\ldots\ldots \quad (7)$$

This is more accurate than equation (**3**).

The exact equation for m when R_l is infinite is found with the aid of Figure 1 to be

$$m = \frac{X \,.\, m_2}{(X+R+R_1)} \left\{ 1 + \left[\frac{R_1^2}{R_2(X+R+R_1)+R_1(X+R)} \right] \right\} \left\{ 1 + \frac{R_1}{(R_1+R_2)-[R_1^2/(X+R+R_1)]} \cdot \frac{X \,.\, m_2}{(X+R+R_1)} \right\}^{-1}.$$

When either $R_1 \ll (X+R)$ or $R_1 \ll R_2$, but especially when both these conditions hold, the terms in square brackets may be neglected. The exact equation then reduces to equation (**4**).

Equations (**4**), (**6**), and (**7**) summarize the principal properties of the amplifier with infinite load resistance for steady D.C. A property which should, perhaps, be emphasized is that the equivalent input e.m.f. representing any residual output p.d. which exists when the input terminals are short-circuited is independent of m_2 and $(1-\beta m_2)$. Such a residual p.d. arises, for example, if the rest position of the galvanometer movement is incorrectly adjusted or if a thermal e.m.f. exists in the galvanometer circuit.

It is not difficult to show that, if R_l is finite, and when $R_1 \ll X$, $R_1 \ll R_2$ and $m \gg 1$, the same results hold provided m_2' is substituted for m_2, where

$$m_2' = \frac{R_l m_2}{R_l + R_1 + R_2}.$$

IV. ANALYSIS FOR GALVANOMETER COIL IN MOTION

(*a*) *The Equation of Motion of the Galvanometer Movement*

We may confine ourselves to the degree of approximation represented by the equivalent circuit of Figure 2. In the absence of mechanical vibration, terms contributing to the torque equation for the galvanometer movement arise from the restoring influence of the suspension, the inertia of the movement, the applied e.m.f., the induced e.m.f. responsible for damping (assumed to be wholly electrical), and the feedback e.m.f.

If the torque on the galvanometer movement per unit e.m.f. applied between the galvanometer terminals is denoted by a, and the galvanometer deflection per unit applied e.m.f. (θ/V_2) is denoted by b, then the restoration coefficient μ for the suspension is given by $\mu=a/b$. For $R_1 \ll X$ the torque due to feedback, L_f, is given by the expression

$$L_f=a\frac{X}{X+R+R_1}\beta_0 V_0\,,$$

that is

$$L_f=\frac{X}{X+R+R_1}\beta_0 m_2\mu\theta.$$

The equation of motion is

$$I\ddot{\theta}+p\frac{R_g}{R_g+\dfrac{R_s(R+R_1)}{R_s+R+R_1}}\dot{\theta}+\left[1-\frac{X}{X+R+R_1}\beta_0 m_2\right]\mu\theta=\frac{X}{X+R+R_1}aV,$$

in which I is the moment of inertia of the movement, and p is a constant. The equation may be written

$$I\ddot{\theta}+D\dot{\theta}+(1-\beta_1 m_2)\mu\theta=MV. \quad \ldots\ldots\ldots\ldots \quad (8)$$

The condition for critical damping is

$$D^2=4(1-\beta_1 m_2)\mu I, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (9)$$

and the natural frequency when $D=0$ is

$$\omega=\sqrt{1-\beta_1 m_2}\sqrt{\mu/I}. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (10)$$

The amplitude solution for an applied sinusoidal e.m.f., V, of angular frequency ω, is

$$\left|\frac{\theta}{V}\right|=\frac{M}{[\{(1-\beta_1 m_2)\mu-\omega^2 I\}^2+\omega^2 D^2]^{\frac{1}{2}}}. \quad \ldots\ldots\ldots\ldots \quad (11)$$

The phase difference between θ and V need not be discussed.

(*b*) *The Effect of Source Resistance on Natural Frequency and Damping*

Equation (**10**) shows that for $D=0$, $R=0$, and R_1 small, feedback results in an increase in the natural angular frequency by a factor of approximately $\sqrt{1-\beta_0 m_2}$. As the source resistance is increased the effect of feedback becomes less, the natural frequency then approaching that of the isolated galvanometer. However, an increase of the source resistance produces a marked change in the damping, which requires some further discussion.

Since in most practical cases the amplifier is operated with the galvanometer motion critically damped or over-damped, it is convenient to express the damping in terms of the dimensionless quantity α, defined by the expression

$$\alpha=\frac{D^2}{4(1-\beta_1 m_2)\mu I}.$$

When α is less than, equal to, or greater than unity the damping is respectively less than critical, critical, or greater than critical.

Apart from small effects consequent upon the finite value of R_1 an increase in the source resistance affects the damping in two ways, namely, the effect of feedback is decreased, resulting in greater damping, and the total resistance in series with the galvanometer coil is increased, diminishing the damping. The two effects thus partially compensate one another and under certain conditions the damping, i.e. α, goes through a minimum as R increases.

As R_1 occurs in association with R in the equation of motion and hence in the expression for α it may for present purposes be regarded as included in R and thus be omitted from the analysis. On replacing $(1-\beta_1 m_2)$ by the approximate expression $\beta_1 m_2$, and substituting $(R_s R_g)/(R_s+R_g)$ for X, we have for the damping

$$\alpha \doteqdot \frac{p^2 R_g{}^2}{4\left(R_g+\dfrac{R_s R}{R_s+R}\right)^2} \cdot \frac{R_s R_g+R(R_s+R_g)}{R_s R_g} \cdot \frac{1}{\beta_0 m_2 \mu I}.$$

In order that the damping should always increase with increasing R, α must have no turning point. A minimum of damping occurs when

$$R=\frac{R_s(R_s-R_g)}{(R_s+R_g)}.$$

If $R_s<R_g$ no minimum occurs in practice, since R cannot be negative, and when $R_s=R_g$ the minimum occurs at zero source resistance. If the motion is adjusted to be critically damped at $R=0$ when $R_s=R_g$, increase in source resistance can only increase the degree of damping. This is a convenient practical adjustment to ensure that the motion will remain non-oscillatory. As may be seen from equation (**9**), in order to achieve critical damping for $R_s=R_g$ and $R=0$, a galvanometer is required which, when isolated, has a very large value of critical damping resistance.

It will be clear from these considerations that scope exists for the development of special galvanometers for use in amplifiers. Such galvanometers should have a high value of critical damping resistance and a very high sensitivity. A galvanometer feedback amplifier differs from an ordinary galvanometer in that the damping increases with increasing source resistance. Except for small values of R the motion is over-damped, but objectionable sluggishness does not occur unless R is very large because of the considerable reduction in the period of the galvanometer resulting from the use of feedback.

V. Response to Alternating Electromotive Forces

From the amplitude solution of the equation of motion (equation (**11**)) it is found by differentiation that the turning points of $\left|\frac{\theta}{V}\right|$ occur at $\omega=0$ and at $\omega=+\sqrt{(2\mu_1 I-D^2)/(2I^2)}$, where μ_1 has been used to represent $(1-\beta_1 m_2)\mu$. The response curve thus has no real maximum unless $2\mu_1 I>D^2$. The condition for critical damping, on the other hand, is $4\mu_1 I=D^2$. If the motion is critically damped the expression for $\left|\frac{\theta}{V}\right|$ (equation (**11**)) becomes

$$\left|\frac{\theta}{V}\right|=\frac{M}{\mu_1[1+(\omega^2/\omega_{00}{}^2)]}, \quad \ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{12})$$

where $\omega_{00}^2=(1-\beta_1 m_2)(\mu/I)$. For no feedback, $\beta_1=0$, and

$$\left|\frac{\theta}{V}\right|=\frac{M}{\mu[1+(\omega^2/\omega_0^2)]}, \quad \dots\dots\dots\dots\dots\dots \quad (\mathbf{13})$$

where $\omega_0^2=\mu/I$. Apart from the expected gain factor $(1-\beta_1 m_2)$, the effect of feedback is to expand the frequency response curve along the ω-axis by the ratio $\sqrt{1-\beta_1 m_2}:1$. If the system is damped to any extent other than critically, this relationship will still apply if the damping, α, is the same with and without feedback. The graph of equation (**13**) is shown in Figure 3.

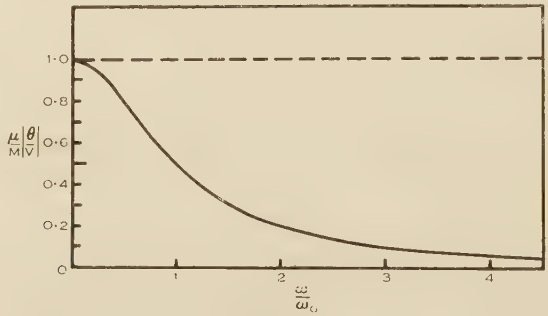


Fig. 3.—The response to alternating e.m.f.'s in the absence of feedback (equation (**13**)).

From the above it may therefore be concluded that the only effect of feedback on the input to noise ratio for electrical noise is due to the greater band of frequencies which can be amplified with the feedback system.

VI. Response to Mechanical Vibration

The response of the system to rotational vibration about the galvanometer suspension as the axis will now be considered. Let θ be an angle measured between lines fixed relative to the galvanometer coil and the main body of the instrument (in particular the arrangement of the light source and phototube), φ an angle which represents the position of the instrument in space, and ζ an angle between lines fixed relative to the instrument frame (including the light source and phototube) and the galvanometer frame and magnet. θ and ζ are assumed to be so selected that they are each zero when the galvanometer coil is in its null output position for zero applied e.m.f.

For zero applied e.m.f. the equation of motion of the galvanometer coil is then

$$I(\ddot{\theta}+\ddot{\varphi})+D(\dot{\theta}+\dot{\zeta})+(1-\beta_1 m_2)\mu\theta+\mu\zeta=0.$$

It is convenient to discuss separately the vibration of the galvanometer relative to the rest of the instrument, assumed fixed in space, and the vibration of the instrument as a whole, the galvanometer frame being assumed rigid with respect to the rest of the instrument. Representing φ by ψ sin $\omega_1 t$ and ζ by η sin $\omega_2 t$, the equation of motion for the first case, namely

$$I\ddot{\theta}+D\dot{\theta}+\mu_1\theta=-\eta(D\omega_2\cos\omega_2 t+\mu\sin\omega_2 t)$$

is obtained on putting $\psi=0$. For critical damping the amplitude solution is

$$\left|\frac{\theta}{\eta}\right|=\frac{\sqrt{1+4(1-\beta_1 m_2)^2\cdot(\omega_2^2/\omega_{00}^2)}}{(1-\beta_1 m_2)[1+(\omega_2^2/\omega_{00}^2)]}, \quad \ldots\ldots\ldots \quad (\mathbf{14})$$

where, as before, $\omega_{00}^2=\mu_1/I$.

In the second case ($\eta=0$) the equation of motion becomes

$$I\ddot{\theta}+D\dot{\theta}+\mu_1\theta=\psi I\omega_1^2\sin\omega_1 t$$

and the amplitude solution for critical damping is

$$\left|\frac{\theta}{\psi}\right|=\frac{\omega_1^2/\omega_{00}^2}{1+(\omega_1^2/\omega_{00}^2)}. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{15})$$

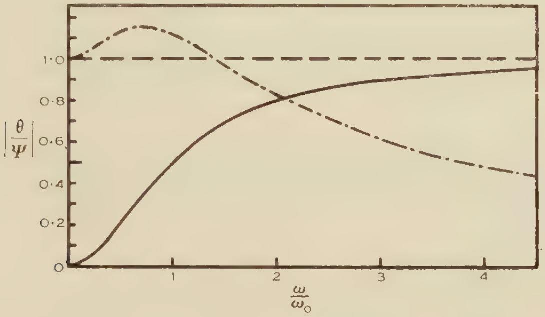


Fig. 4.—The response to mechanical vibration in the absence of feedback.

——— Vibration of the instrument as a whole (equation (**15′**)).

—·—·— Vibration of the galvanometer frame relative to the rest of the instrument (equation (**14′**)).

The corresponding expressions when there is no feedback, ($\beta_1=0$), are

$$\left|\frac{\theta}{\eta}\right|=\frac{\sqrt{1+(4\omega_2^2/\omega_0^2)}}{1+(\omega_2^2/\omega_0^2)}, \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{14'})$$

and

$$\left|\frac{\theta}{\psi}\right|=\frac{\omega_1^2/\omega_0^2}{1+(\omega_1^2/\omega_0^2)}, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{15'})$$

where $\omega_0^2=\mu/I$.

Graphs of equations (**14′**) and (**15′**) are shown in Figure 4. It is clear from comparison with Figure 3 that a galvanometer is sensitive to a much greater range of frequencies for rotational vibration of the galvanometer relative to the projector and scale system than for an applied sinusoidal p.d.

The conditions represented by equation (**15′**) would only arise in galvanometers with an inbuilt light and scale system. Here the sensitivity to rotational vibration of the whole instrument increases with increasing frequency of vibration, the sensitivity in this case being the complement with respect to unity of the sensitivity to applied e.m.f. plotted in Figure 3.

It follows from equation (**14**) that as ω_2/ω_{00} is increased from $\omega_2=0$, when $\left|\frac{\theta}{\eta}\right|=\frac{1}{1-\beta_1 m_2}$, the value of $\left|\frac{\theta}{\eta}\right|$ increases to a maximum of 1 (approximately) at $\omega_2/\omega_{00}=1$ and then progressively decreases. For values of ω_2/ω_{00} large compared with unity, $\left|\frac{\theta}{\eta}\right|$ approximates to $2(\omega_{00}/\omega_2)$. It is thus clearly desirable that ω_2 should be either much greater or much less than ω_{00}. Since $\omega_{00}=\sqrt{1-\beta_1 m_2}\,.\,\omega_0$, where ω_0 is the natural frequency of the galvanometer coil, ω_{00} will have a value of about 20 c/s. for a galvanometer having a period of 2 sec. if the value of $(1-\beta_1 m_2)$ is about 1500.

For vibration of the instrument as a whole relative to a fixed direction in space, as represented by equation (**15**), it is clear that $\left|\frac{\theta}{\psi}\right|$ is zero for $\omega_1/\omega_{00}=0$ and increases, without turning point, to approach unity as ω_1/ω_{00} becomes large. It is therefore advantageous, other things being equal, for ω_1 to be made as small as possible.

From equations (**12**), (**14**), and (**15**) it follows that, to obtain the e.m.f. equivalent of the vibrational disturbances discussed, the angular disturbance must be multiplied by

$$\frac{\mu}{M}\sqrt{1+4(1-\beta_1 m_2)^2(\omega_2{}^2/\omega_{00}{}^2)}, \qquad \text{for the case } \psi=0$$

and

$$\frac{\mu(1-\beta_1 m_2)(\omega_1{}^2/\omega_{00}{}^2)}{M}, \qquad \text{for the case } \eta=0.$$

From equations (**13**) and (**14′**) the e.m.f. equivalent of a vibrational disturbance for no feedback for $\psi=0$ and critical damping is the same as with feedback for static conditions $(\omega_2=0)$; for $\eta=0$ the disturbances are in each case zero for $\omega_1=0$.

For a given amplitude of vibration ψ or η the sensitivity to rotational vibration of the system with negative feedback, expressed in terms of an equivalent input e.m.f., increases very rapidly with increase in ω_1 or ω_2 as compared with the corresponding sensitivity of the galvanometer without negative feedback. The use of feedback in a galvanometer amplifier thus not only fails to improve the input to noise ratio, but, as far as rotational vibrations (about the suspension as axis) are concerned, renders it considerably worse. Translational vibrations and rotational vibrations about axes perpendicular to the suspension do not result in the moving coil cutting the field to any extent, and, since they also have little influence on the system through the optical system to the phototube, the amplifier is almost insensitive to them.

VII. References

Frankenhaeuser, B., and MacDonald, D. K. C. (1949).—*J. Sci. Instrum.* **26**: 145.

Gall, D. C. (1942).—*J. Instn. Elect. Engrs.* **89** (2): 434.

Hill, A. V. (1948).—*J. Sci. Instrum.* **25**: 225.

Preston, J. S. (1946).—*J. Sci. Instrum.* **23**: 173.

Wylie, R. G. (1951).—A galvanometer amplifier for the accurate amplification of small D.C. electromotive forces. C.S.I.R.O. Div. Phys., Rep. PA-5.

THE PHOTOVOLTAIC EFFECT IN NATURAL LEAD SULPHIDE

By R. LAWRANCE*

[*Manuscript received June 5, 1951*]

Summary

Australian samples of natural lead sulphide have been used as radiation detectors in the near infra-red. The nature of the photovoltaic effect has been investigated and the sensitivity of some cells measured.

The electrical properties of bulk samples have been measured. The results have been compared with similar measurements given by other authors for thin lead sulphide films, and a close correlation has been established between them.

I. THE GENERATION OF ELECTRICAL POWER BY LEAD SULPHIDE CRYSTALS

(*a*) *The Point Contact Cell* (*Bose* 1904)

When the area immediately surrounding the contact surface between a metal and a lead sulphide crystal is illuminated by radiation between the wavelengths of 0·5 and 4 μ, three separate electrical effects may be present, namely, thermoelectric, photovoltaic, and photoconductive effects. The first two combine to produce an electrical potential difference in the vicinity of the contact, and are only separable by their difference in time constant or by a difference in polarity. The third effect, if present, shows itself as a change in conductivity.

(*b*) *Experimental*

For the following experiments, cells similar to crystal detectors were used. Pieces of galena were mounted in Wood's alloy and thin wires of Mo, Cu, or Ni were held under light pressure against the surface of the galena to form " point contacts " (see Fig. 1). The cells were mounted on a mechanical stage capable of fine movement in three directions at right angles, and rigidly supporting a Schwarzschild objective ($f/1$). The source of radiation was a 12 V. 24 W. lamp with a straight tungsten-coil filament 4 mm. long by 1 mm. in diameter. An image of this filament was formed on the surface of the galena at a linear reduction of 40 : 1.

In the following discussion of results, the experimental arrangements as set out above will be referred to by the term " working conditions ".

(*c*) *Results*

Under working conditions, the cells were found to yield, when continuously illuminated, a potential difference which was in most cases in the range 20–100 μV. When the radiation falling on the cells was chopped at 9 c/s., and the potential difference across the cells fed into an A.C. amplifier and headphones, in most cases no signal could be detected, although the overall sensitivity of the amplifier was of the order of 10 μV. It was assumed, therefore, that this potential difference under continuous illumination was due to thermoelectric or some other effect of large time constant. Throughout this paper, this potential will be called " *T*-effect ".

---

* Defence Research Laboratories, Department of Supply, Melbourne.

It is possible, by trial and error, to locate small areas of some samples which produce both *T*-effect and another potential which still manifests itself when the radiation is chopped. The magnitude of this residual cell output remains unchanged even when the chopping frequency is increased to at least 2 kc/s. It is therefore certain that this is not a thermoelectric effect, and from other considerations, including previous published work, it is believed that the residual output is associated with barrier layers. This phenomenon will be referred to as " *B*-effect ". It has been found that the effect is present even when spots somewhat removed from the area of contact are illuminated, which indicates that the function of the metal needle is merely to complete the electrical circuit.

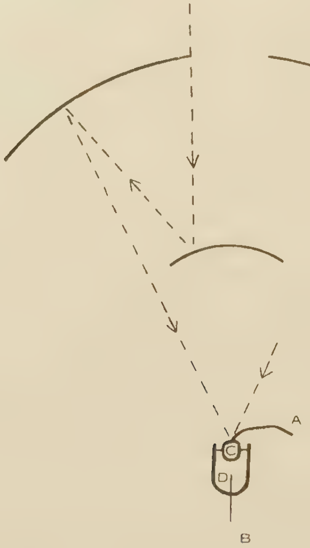


Fig. 1.—Schwarzschild objective and point contact cell.
A, contact ; *B*, electrode ; *C*, crystal ; *D*, Wood's alloy.

Samples which exhibited *B*-effect most strongly were always found to be of the finely crystalline type, though the converse is not true. Well-formed cubes were always found to exhibit little or no *B*-effect. The sources of the samples were :

1. Zinc Corp. Ltd., Broken Hill, New South Wales.
2. Central Drawing Office, Department of Supply (origin unknown).
3. Zinc Corp. Ltd., Broken Hill, New South Wales.
4. Farrell Mining Co. Ltd., Tullah, Tasmania.
5. Broken Hill South Ltd., Broken Hill, New South Wales.

6. Elliot Lead-Central Lead, Lake George Mines Pty. Ltd., Captain's Flat, New South Wales.
7. North Broken Hill Ltd., Broken Hill, New South Wales.
8. Mt. Isa Mines Ltd., Mt. Isa, Queensland.
9. Electrolytic Zinc Co. of Australasia Ltd., Read-Rosebery Mines, Tasmania.
10. Lilydale, Victoria (National Museum, Melbourne).

The behaviour of a number of cells of high output, all from Tullah, Tasmania, is summarized in Table 1. The resistance is that which, as a load, gives maximum power transfer under working conditions. Peak values of voltage, and total radiation from the source (as measured with a thermopile) are used to calculate sensitivity and minimum detectable power at 9 c/s. chopping frequency and 1 c/s. bandwidth (9 c/s. was found to eliminate *T*-effect). Because the cut-off for these cells is about 3 μ, the error introduced by using total radiation is negligible.

TABLE 1

THE SENSITIVITY OF TYPICAL HIGH OUTPUT CELLS

($T=293$ °K.)

Sample (Tullah, Tas.)	A.C. Resistance (ohms at 9 c/s.)	Sensitivity (μV./μW. radiation)	Min. Det. Power* (W. at 1 c/s. bandwith)
1	1500	59·9	$8 \cdot 2 \times 10^{-11}$
2	1500	11·3	$4 \cdot 4 \times 10^{-10}$
3	1500	15·1	$3 \cdot 3 \times 10^{-10}$
4–1	1500	3·7	$1 \cdot 3 \times 10^{-9}$
4–2	13000	14·4	$1 \cdot 0 \times 10^{-8}$
4–5	2000	34·6	$1 \cdot 6 \times 10^{-10}$
4–7	7000	16·6	$6 \cdot 4 \times 10^{-10}$
4–8	6000	20·8	$4 \cdot 8 \times 10^{-10}$

* Minimum detectable power$=2(kRT\delta f)^{\frac{1}{2}}$/sensitivity, where k = Boltzmann's constant, R = resistance, T = absolute temperature, and δf = bandwidth.

Tests were made on some of these cells to determine the D.C. resistance. Because the voltage-current curves for these cells are asymmetrical and non-linear under most conditions, currents during these measurements were limited to the same order as the average A.C. load current generated by the cells under working conditions. At these currents no asymmetry was evident.

The polarities of the *T*- and *B*-effects were determined. For the *B*-effect, this was achieved by sending short pulses of radiation to the cells and studying the signs of the outputs on a cathode-ray oscilloscope. The polarities are given in Table 2, together with thermoelectric and photovoltaic outputs and resistances for comparison.

Insufficient tests have been made to establish conclusively the similarity between the polarity of the two effects, or to point to any definite relationship between their magnitudes.

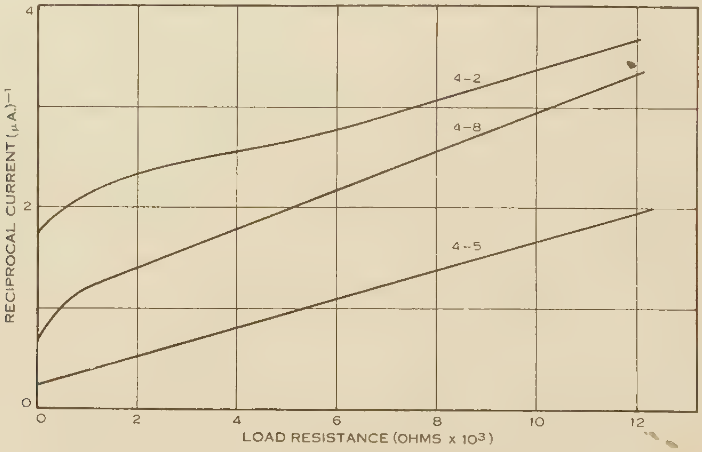


Fig. 2.—Current characteristics of typical sensitive cells.

Power transfer measurements were carried out on one cell with different contact pressures, i.e. different contact resistances, and no variation in the cell resistance (as determined from maximum power transfer) could be observed.

Table 2

A COMPARISON OF FOUR CELLS FROM TULLAH, TASMANIA

Sample	D.C. *T*-Effect* (μV.)	Peak A.C. *B*-Effect (μV.)	*T*-Effect Polarity	*B*-Effect Polarity	D.C. Resistance (ohms)	A.C. Resistance (ohms)
4–1	1273	3280	Contact +	Contact +	1300	1500
4–5	4189	7050	,, +	,, +	1520	2000
4–7	2899	6504	,, +	,, +	5680	7000
4–8	3591	6391	,, —	,, —	Not measured	6000

* *T*-effect was determined by subtracting peak *B*-effect voltage from the total D.C. voltage developed by the cell under steady illumination.

It is clear, then, that the resistance of the cell lies not in the contact but in the crystal. To investigate the nature of this resistance, the load current was measured with changing load resistance. In most cases the relationship between the two is strictly ohmic, though in some cases the current at small load resistances was higher than would be expected from the ohmic relationship (Figs. 2, 3).

The power transfer and noise characteristics of the cells have been plotted in Figures 4 and 5 for cells which are ohmic and non-ohmic respectively. The

" power transfer ratio " curves were taken from actual measurements on the cells, and were found to agree very closely with the theoretical expression P.T.R. $=4rR/(r+R)^2$, where r is the internal resistance of the cell and R the load resistance of the input circuit of the power amplifier. The curve for Johnson noise power ratio has been plotted from the expression J.N.P.R. $=2R/(r+R)$. The ordinates in Figures 4 and 5 have been so adjusted that they are equal to unity for all curves when $R=r$. It is clear that, for the ohmic case (Fig. 4), a reduction by a factor two in the noise to signal ratio can be achieved by reducing the load resistance to as near zero as possible.

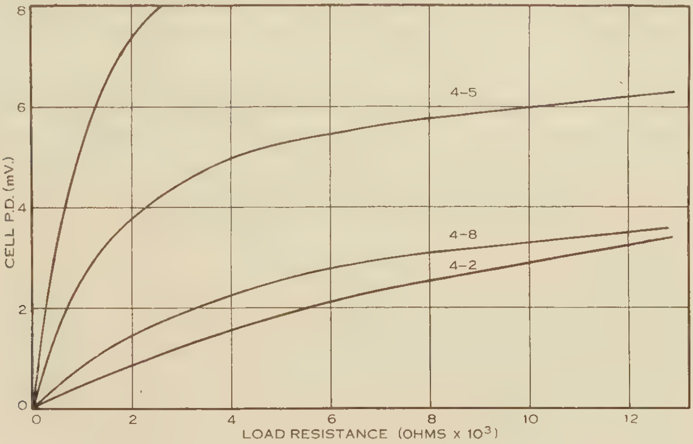


Fig. 3.—Voltage characteristics of typical sensitive cells.

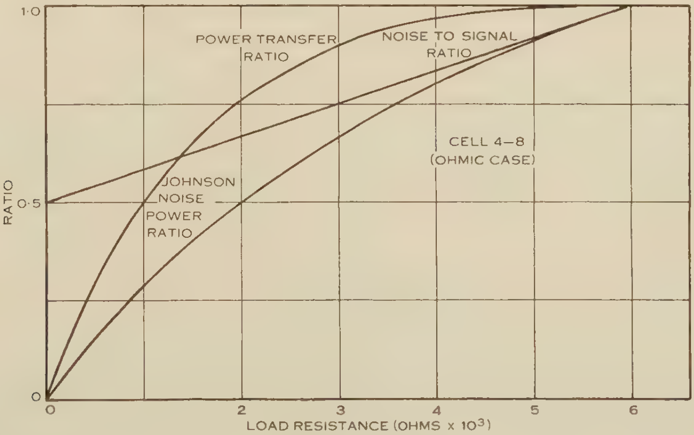


Fig. 4.—The influence of the load resistance on the noise to signal ratio for ohmic cells.

In the non-ohmic case it is possible to improve the noise to signal ratio only slightly by using the matching resistance indicated by the minimum value of this curve.

Response to modulation is extremely fast. Results of the present experiments show no change in output for continuous modulation up to 2·75 kc/s., and no apparent decrease for radiation pulses as short as 10^{-4} sec. At present, lack of suitable apparatus prohibits measurements being made at shorter durations than this, but it has been shown (Dr. F. W. Woltersdorf, personal communication) that for Sardinian samples of lead sulphide, a signal is still present for pulses as short as 10^{-6} sec.

With certain cells it is found that if a secondary source of e.m.f. is introduced in series, a marked increase in output is experienced when the source has the correct polarity. On increasing the magnitude of this e.m.f. above a certain limit, no further increase in output is observed, and the sensitive area is often destroyed. Because the noise to signal ratio increases markedly with introduction of this e.m.f., the effect was not investigated further.

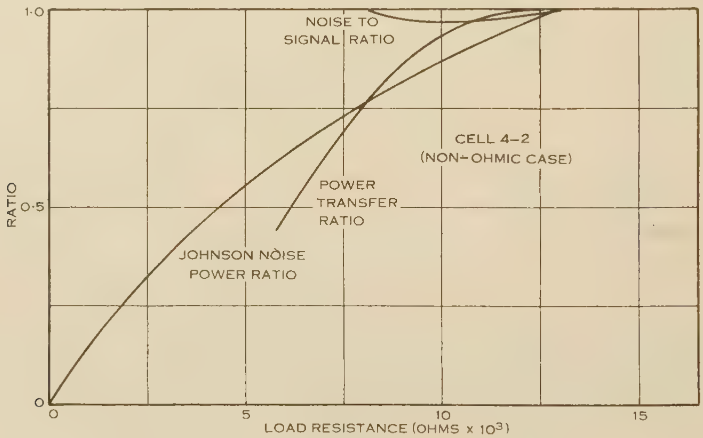


Fig. 5.—The influence of the load resistance on the noise to signal ratio for non-ohmic cells.

II. Electrical Properties of Bulk Samples of Lead Sulphide Crystals

(a) *Thermoelectric Power*

The thermoelectric power of lead sulphide is a quantity which varies over a considerable range of positive and negative values, depending upon the departure from the stoichiometric composition caused both by impurities and by excess of one of the major elements. The range for films is known to be from about −500 to +400 μV./°C. against platinum (Hintenberger 1942 ; Sosnowski, Starkiewicz, and Simpson 1947), and measurements by Eisenmann (1940) and Legrand (1948) established a similar range for natural crystals.

The results of the present experiments, with the exception of those for one sample, are all negative and range from −88 to −400 μV./°C. against copper. The exception is a sample with thermoelectric power of +390 μV./°C.

The apparatus used for these measurements consisted of a pair of spring-loaded copper rods $\frac{1}{8}$ in. in diameter, between which each sample, about 5 mm. long by 2 mm. square, was clamped. A difference in temperature of up to 100 °C. was maintained across the contacts by electrical heating elements wound on the ends of the rods, the temperatures at the faces being measured by a pair of fine constantan wires set into the ends of the rods. A set of switches enabled the e.m.f.'s due to each thermocouple, and that across the copper rods, to be measured in quick succession, from which the thermoelectric output between those temperatures was determined. A number of readings at various temperatures enabled the curves of thermoelectric output against temperature to be drawn and the gradients read off at any desired temperature.

(*b*) *Conductivity*

Conductivity was measured using the same apparatus and samples as above. The samples were heavily silvered on both ends to facilitate good contact. It was found essential to keep the end temperatures the same to eliminate error due to thermoelectric potentials across the crystal. This condition was maintained by gently warming the cooler rod until zero potential was produced across the contacts. A small current was then passed through the sample and the potential drop across the faces measured. During this operation a check was kept upon the end temperatures by reference to the existing thermocouples at the faces.

The results show considerable variation among samples, ranging from 0·2 to 41·2 $\text{ohm}^{-1}\ \text{cm.}^{-1}$. In general, well-formed cubes have a much higher conductivity than crystal aggregates. The values differ widely from values obtained by Hintenberger (1942) for thin films of varying sulphur content, and from the results of other measurements on photoconductive films during sensitization (Sosnowski, Starkiewicz, and Simpson 1947). Eisenmann (1940) and Chasmar (1948) have already reported this conductivity difference between crystals and films.

(*c*) *Hall Constant*

Finally, measurements of Hall constant were made in fields of 3000 gauss. Rectangular samples a few mm. long by about 1 mm. thick were used, with wedge-shaped probes to average out variations in the Hall potential across the edges and reduce thermoelectric drifts due to fluctuating ambient temperature. The results should not be taken as being better than ±10 per cent. due to the inherent difficulties and the low potentials encountered with samples of this thickness.

The results of measurements of Hall constant (R), conductivity (σ), and mobility ($\mu=R\sigma$) are set out in Table 3 together with the results of Legrand (1948). A striking difference in the results presented here is that those of the present experiments give the same sign for Hall constant and thermoelectric power, whereas Legrand's results show opposite signs.

TABLE 3

MEASUREMENTS OF CONDUCTIVITY, HALL CONSTANT, AND MOBILITY

Present Measurements					Legrand's Measurements		
Sample	Thermo-electric Power (μV./°C.)	Hall Constant (cm.^{3}/C.)	Specific Con-ductivity (ohm^{-1} cm.$^{-1}$)	Mobility (cm.^{2}/V. sec.)	Sample	Thermo-electric Power (μV./°C.)	Hall Constant (cm.^{3}/C.)
1	—254	— 2·7	13·3	30·5	Sardinia	—519	+ 16
2	—375	—11·7	0·2	2·3	,,	—514	+195
3	—233	— 2·0	41·2	82·4	,,	—492	+198
4	—361	—13·3	0·58	7·6	,,	—489	
5	—235	— 2·3	4·0	9·2	,,	—515	
6	—117	Not measured	0·22	—	,,	—425	
7	—312	— 7·6	17·3	131·5	,,	—500	
8	—218	— 2·1	23·9	50·2	Ems	+427	— 83
9	—347	— 8·4	13·8	116	Joplin	+300	
10	+390	+498	0·26	100	Freiberg	+110	
					Sardinia	+365	— 67
					,,	+372	— 68

III. DISCUSSION

The electrical properties of lead sulphide films, and in particular those properties which have already been referred to in the present paper, have been discussed by various authors in terms of the excess and defect conduction processes. It has been found that a definite relationship exists between the photoelectric properties of the films and the measured values of their thermoelectric power, Hall constant, and conductivity, which are the physical properties used in determining the stoichiometric composition of the material. There seems no reason to believe that these findings should not apply equally well to natural crystals; the latter are formed in the presence of other elements, such as oxygen, which are deliberately introduced in the manufacture of the synthetic films (Hintenberger 1946; Sosnowski, Starkiewicz, and Simpson 1947; Frank and Raithel 1949). The algebraic signs of the results of Hall constant and thermoelectric power for Australian natural crystals, as set out in Table 3, show that, with one exception, the excess conduction process obtains in these crystals.

In Figure 6 it is evident that the relationship between Hall constant and thermoelectric power is the same for these crystals as for films in which excess conductivity predominates (Hintenberger 1942).

On the other hand, the figures for conductivity of crystals given in Table 3 are 100 to 1000 times greater than Hintenberger's (1942) figures for lead sulphide films. This is illustrated graphically in Figure 7. This figure shows, however, that the high values of conductivity σ (where $\sigma = ne\mu$) for crystals are accompanied by a correspondingly high mobility μ of the carriers, each of charge e, and that

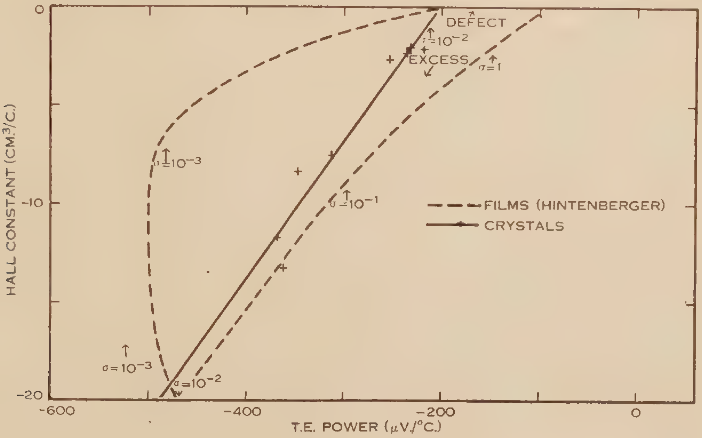


Fig. 6.—The Hall constant and thermoelectric power of lead sulphide crystals compared with that of synthetic films in the region of excess conductivity.

the number of carriers per unit volume, n, as given by the slope of the curve, is the same for both crystals and films over a range of conductivity from 10^{-3} to 10^{2} ohm^{-1} cm.$^{-1}$.

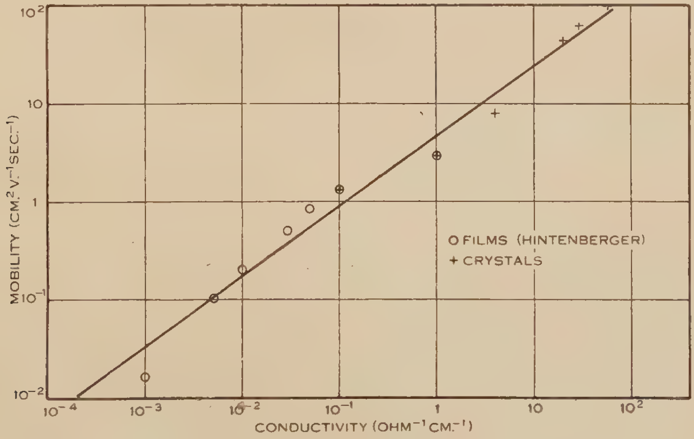


Fig. 7.—The relation between mobility and conductivity for lead sulphide crystals and synthetic films with varying sulphur content.

The general trend of photoelectric sensitivity of crystals has been found to be the same as that reported by Sosnowski, Starkiewicz, and Simpson (1947) for films, i.e. the photoelectric sensitivity increases as the conductivity decreases.

The facts given above show a close relationship between lead sulphide films and crystals, and provide evidence justifying the assumption already made, that the electric and photoelectric properties of crystals of semiconducting substances such as lead sulphide can be understood by applying the concepts of semiconductive processes already successfully established for thin films of these substances.

It is hoped that these concepts may be applied to photoelectric detectors made from natural lead sulphide crystals, to assess how closely such detectors approach the maximum possible sensitivity.

In order to establish numerically the relationship between photoelectric sensitivity and conductivity, it will be necessary to carry out measurements on a great number of sensitive spots and submit the results to a statistical analysis to allow for variations in output from spot to spot on the same crystal.

Further work will also be necessary to establish whether oxygen or sulphur centres which, in the light of previous publications (Hintenberger 1946; Sosnowski, Starkiewicz, and Simpson 1947) are the most probable elements creating excess conductivity, are built into the lattice thus accounting for the values of the various properties reported in this paper.

IV. Acknowledgments

The author wishes to thank Dr. F. W. Woltersdorf for his guidance and advice throughout the course of the work; Mr. R. Galbreath for assistance in developing the apparatus used; the workshop staff for its construction; and the Bureau of Mineral Resources, Geology and Geophysics, Department of National Development, for assistance in obtaining samples. The author is indebted to the Chief Scientist, Department of Supply, for permission to publish this paper.

V. References

Bose, J. C. (1904).—Detector for electrical disturbances. U.S. Pat. 755,840 (March 29).

Chasmar, R. P. (1948).—High frequency characteristics of lead sulphide and lead selenide layers. *Nature* **161**: 281-2.

Eisenmann, L. (1940).—*Ann. Phys.* **38**: 121. (FIAT Review of German Science 1939-46: "Physics of Solids." Pt. 2, p. 81.)

Frank, K., and Raithel, K. (1949).—Uber ein Verfahren zur Herstellung lichtelektrisch wirksamer Schichten von sehr hohem Widerstand mit PbS als ultrarotempfindlichen Halbleiter. *Z. Phys.* **126**: 377-82.

Hintenberger, H. (1942).—Uber die elektrischen Eigenschaften des Bleisulfids. *Z. Phys.* **119**: 1.

Hintenberger, H. (1946).—Uber den Einfluss von Gasen, besonders von Sauerstoffspuren, auf die elektrischen Eigenschaften von aufgedampften PbS-Schichten. *Z. Naturforsch.* **1**: 13-7.

Legrand, K. (1948).—Uber den Zusammenhang von Thermokraft und Hall-Effect mit dem lichtelektrischen Effect beim naturlichen Bleiglanz. *Z. Phys.* **124**: 219-24.

Sosnowski, L., Starkiewicz, J., and Simpson, O. (1947).—Lead sulphide photoconductive cells. *Nature* **159**: 818-9.

THE INFRA-RED SPECTRA OF PORPHYRINS AND THEIR IRON COMPLEXES

By J. E. Falk* and J. B. Willis†

[*Manuscript received July* 12, 1951]

Summary

Infra-red absorption spectra between 4000 and 670 cm.$^{-1}$ have been measured for a number of porphyrins and their iron complexes (haems). These spectra should be of use in the identification and purification of porphyrins, and enable position isomers to be distinguished. Characteristic absorption frequencies for the carbonyl group in different environments have been found.

I. Introduction

The present study was begun at the time when the structure of haem-*a* and the related iron-free porphyrin-*a* (Rawlinson and Hale 1949) was also being sought by comparison of their visible spectra with those of pyrrole pigments of known structure (Lemberg and Falk 1951). It was evident (Lemberg *et al*. 1949) that a carbonyl side-chain was present on the porphyrin nucleus. For this reason the infra-red spectra of a number of natural and synthetic porphyrins and haems with and without such side-chains were measured, in the hope that sufficient pure porphyrin-*a* might become available to compare with them. This hope, however, has not yet been realized, but the infra-red spectral study has been extended to include most of the natural porphyrins and their haems.

It is considered worthwhile to publish the results obtained, as infra-red spectroscopy offers a valuable means of identifying and characterizing compounds of this nature, and can often be used to distinguish between closely related position isomers. The position isomers coproporphyrin I and III have both been synthesized by H. Fischer and are well-characterized substances; we have confirmed the differences in their infra-red spectra reported by Gray, Neuberger, and Sneath (1950). The characterization of the isomeric uroporphyrins I and III, however, has been the subject of some controversy, and neither has yet been synthesized. The uroporphyrin I (sample α) and uroporphyrin III (sample β) are thought to be the pure isomers, on the basis of their chromatographic homogeneity and their decarboxylation to the corresponding coproporphyrins I and III respectively (Rimington 1939; Nicholas and Rimington 1951). Comparison with the spectrum of a typical "Waldenström ester" supports previous evidence that the latter is a mixture.

---

* Institute of Medical Research, Royal North Shore Hospital, Sydney; and Department of Chemical Pathology, University College Hospital Medical School, London, W.C.1.

† Division of Industrial Chemistry, C.S.I.R.O., Melbourne.

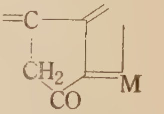
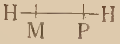
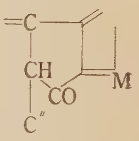
TABLE 1

FORMULAE OF COMPOUNDS

Compound No.	Name	State	Formula*			
1	Octamethylporphin		M	M	M	M
			M	M	M	M
2	Aetioporphyrin I		M	E	M	E
			E	M	E	M
3	Deuteroporphyrin	Dimethyl ester	M	H	M	H
4	Deuterohaem	,, ,,	M	P	P	M
5	Mesoporphyrin	,, ,,	M	E	M	E
6	Mesohaem	,, ,,	M	P	P	M
7	Protoporphyrin	,, ,,	M	V	M	V
8	Protohaem	,, ,,	M	P	P	M
9	Protoporphyrin	Free acid	M	V	M	V
10	Protohaem	,, ,,	M	P′	P′	M
11	Haematoporphyrin(a)	Dimethyl ester	M	B	M	B
12	Haematohaem	,, ,,	M	P	P	M
13	Tetramethylhaematoporphyrin(b)	,, ,,	M	Bm	M	Bm
			M	P	P	M
14	Coproporphyrin I (α and β)	Tetramethyl ester	M	P	M	P
			P	M	P	M
15	Coproporphyrin III	,, ,,	M	P	M	P
			M	P	P	M
16*a*	2-Monoformyldeuteroporphyrin	Dimethyl ester	M	F	M	H
			M	P	P	M
16*b*	4-Monoformyldeuteroporphyrin	,, ,,	M	H	M	F
17	4-Monoformyldeuterohaem(c) (d)	,, ,,	M	P	P	M
18	Diformyldeuteroporphyrin	,, ,,	M	F	M	F
19	Diformyldeuterohaem	,, ,,	M	P	P	M
20	Monoacetyldeuteroporphyrin	,, ,,	M	H	M	Ac
21	Monoacetyldeuterohaem	,, ,,	M	P	P	M
22	Diacetyldeuteroporphyrin	,, ,,	M	Ac	M	Ac
23	Diacetyldeuterohaem	,, ,,	M	P	P	M
24	Chlorocruoroporphyrin	,, ,,	M	F	M	V
			M	P	P	M

TABLE 1 (*Continued*)

FORMULAE OF COMPOUNDS

Compound No.	Name	State	Formula*			
25	Rhodoporphyrin	Free acid	M	E	M	E
			M	P′	C′	M
26	Rhodohaem(e)	Dimethyl ester	M	E	M	E
			M	P	C″	M
27	Uroporphyrin I(f)	Octacarboxylic acid	A′	P′	A′	P′
			P′	A′	P′	A′
28	Uroporphyrin I (α and β)(f)	Octamethyl ester	A	P	A	P
29	Urohaem I	,, ,,	P	A	P	A
30	Uroporphyrin III (β)(f)	,, ,,	A	P	A	P
			A	P	P	A
31	Uroporphyrin (γ)(g)	,, ,,				
32	Phylloerythrin porphyrin	Monomethyl ester	M	E	M	E
33	Phylloerythrin haem	,, ,,	M	P		
			M	V	R†	E
34	Methylphaeophorbide-*a*					
35	Methylphaeophorbide-*b*(c)					

* Key to Table 1.—H = —H ; M = $—CH_3$; E = $—CH_2CH_3$; V = $—CH{=}CH_2$; F = —CHO ; Ac = $—CO–CH_3$; B = $—CH(OH)CH_3$; Bm = $—CH(OCH_3)CH_3$; C′ = —COOH ; C″ = $—COOCH_3$; A′ = $—CH_2COOH$; A = $—CH_2COOCH_3$; P′ = $—CH_2CH_2COOH$; P = $—CH_2CH_2COOCH_3$.

† Where R = M in (34) and F in (35).

(a) Found to contain large quantities of the tetramethyl compound.

(b) The infra-red spectrum of this compound has recently been published by Paul (1951), and resembles the present spectrum.

(c) Insufficient quantities available for satisfactory recording of the spectrum in the solid state. The spectrum in the 1600–1800 cm.$^{-1}$ region was obtained, however, in chloroform solution.

(d) Probably a mixture of the 2- and 4-formyl isomers.

(e) Somewhat contaminated with the free acid.

(f) Constitution not proved by synthesis.

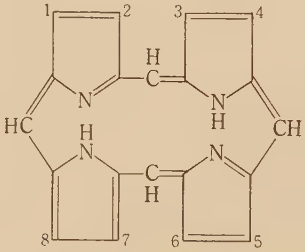
(g) Shown to be a mixture (see Section V).

It was also found possible to establish characteristic absorption frequencies for the carbonyl group in different environments, and this should help in the assignment of structures to new compounds.

II. PREPARATION OF MATERIALS

For ease of reference the compounds have been numbered with Arabic numerals to avoid confusion with the Roman numerals used in the conventional manner to distinguish position isomers.

The basic nucleus of all the compounds is porphin



and Table 1 shows the formulae of the individual compounds by indicating the substituents in the eight positions in this nucleus. Thus, for instance, protoporphyrin dimethyl ester, which is dimethyl 1,3,5,8-tetramethyl-2,4-divinylporphin-6,7-dipropionate, is conveniently written

M	V	M	V
M	P	P	M

where M, P, and V have the meanings shown in the key to Table 1.

(i) *Melting Points.*—These were determined by means of a Gallenkamp electric hot-stage melting-point apparatus (Mason 1925), and were corrected.

(ii) *Absorption Spectra.*—Absorption spectra in the visible region for some of the compounds were measured in dioxane solution with a Beckman photoelectric spectrophotometer, and the absorption bands are summarized in Table 2. The spectra of compounds 16, 18, 20, 22, 24, and 32 were measured by Lemberg and Falk (1951).

Haemins or their esters were prepared from the corresponding pure porphyrins or esters by the ferrous acetate-acetic acid method of Warburg and Negelein (1932).

Source of Porphyrins

(*1*) *Octamethyl porphin.* No m.p. Supplied by Dr. R. Lemberg.

(*2*) *Aetioporphyrin I.* No m.p. Supplied from the collection of the late Professor H. Fischer by Professor S. Goldschmidt.

(*3*) *Deuteroporphyrin dimethyl ester.* Deuterohaemin, prepared from haemin by the resorcinol fusion method (Schumm 1928), was converted by the method of

Grinstein (1947) to deuteroporphyrin dimethyl ester. After saponification, this was purified by ether-HCl fractionation, re-esterified, and after three crystallizations from chloroform-methanol its m.p. was 224·5 °C.

Table 2

ABSORPTION SPECTRA OF PORPHYRINS IN THE VISIBLE REGION

Compound		Positions of Maxima (mμ)				Intensities, relative to Band IV			
		I	II	III	IV	I	II	III	IV
2	(*a*)	621	567	528	496	0·390	0·462	0·712	1·0
	(*b*)	621	566	528	496	0·381	0·438	0·699	1·0
3	(*a*)	618	565	525	495	0·274	0·430	0·536	1·0
	(*e*)	618	565	525	495	0·272	0·426	0·538	1·0
5	(*a*)	620	567	528	496	0·376	0·459	0·690	1·0
	(*c*)	620	567	528	496	0·380	0·462	0·690	1·0
7	(*a*)	630	575	537	502	0·380	0·461	0·793	1·0
	(*b*)	630	575	537	503	0·381	0·463	0·790	1·0
25	(*a*)	632	573	545	506	0·184	0·699	1·270	1·0
	(*b*)	632	573	545	506	0·182	0·706	1·290	1·0
32	(*a*)	634	582	557	518	0·224	1·060	1·469	1·0
	(*c*)	634	581·5	557·5	517·5	0·218	1·000	1·517	1·0
34	(*a*)	667	610	535	506	4·720	0·725	0·848	1·0
	(*c*)	666	610	535	506	4·770	0·706	0·842	1·0
35	(*a*)	653	600	556	526	2·550	0·653	0·641	1·0
	(*d*)	652·5	600	555	525	2·660	0·660	0·650	1·0

(*a*) This paper.
(*b*) Stern and Wenderlein (1934).
(*c*) Stern and Wenderlein (1935*a*).
(*d*) Stern and Wenderlein (1935*b*).
(*e*) Stern and Wenderlein (1935*c*).

(*5*) *Mesoporphyrin dimethyl ester.* Mesoporphyrin was prepared from haemin by the method of Fischer and Röse (1913). The porphyrin was esterified with cold HCl-saturated methanol and crystallized three times from chloroform-methanol, m.p. 216 °C.

(*7*) *Protoporphyrin dimethyl ester.* This was prepared from haemin by the method of Grinstein (1947), m.p. 231 °C.

(*9*) *Protoporphyrin free dicarboxylic acid.* Prepared by saponification of the pure ester, and crystallized from pyridine and then from ether; it has no m.p.

(*14*) *Coproporphyrin I tetramethyl ester.* Sample α was prepared from calf meconium, m.p. 250–251 °C. Sample β was synthetic material from the Fischer collection, m.p. 252 °C.

(*15*) *Coproporphyrin III tetramethyl ester.* Prepared from the culture filtrate from *Corynebacterium diphtheriae*, m.p. 155–157 °C., remelt 181–182 °C. (cf. Gray and Holt 1948).

(*16a*) and (*16b*) were prepared by Lemberg and Parker (unpublished data).

(*17*), (*18*), (*20*), and (*22*) were prepared as described by Lemberg and Falk (1951).

(*24*) *Chlorocruoroporphyrin*, m.p. 276–278 °C., was prepared synthetically from protoporphyrin (Lemberg and Parker, unpublished data).

(*25*) *Rhodoporphyrin free acid.* Supplied from the Fischer collection, m.p. 264 °C.

(*27*) *Uroporphyrin I octacarboxylic acid.* Supplied by Dr. R. Lemberg.

(*28*) *Uroporphyrin I octamethyl ester.* Sample α was isolated from bovine congenital porphyria bone (Rimington 1936), m.p. 293 °C. Sample β was isolated from the urine of a human congenital porphyria patient, m.p. 293 °C. (Rimington and Miles 1951).

(*30*) *Uroporphyrin III octamethyl ester.* Sample β was prepared from turacin ; from the results of a study of this material, m.p. 264 °C., Nicholas and Rimington (1951) consider that it is probably pure uroporphyrin III, free from uroporphyrin I.

(*31*) *Uroporphyrin.* Sample γ was isolated from the urine of a human patient suffering from porphyria cutanea tarda, m.p. 256–260 °C. This is considered to be a typical " Waldenström ester " (Waldenström 1937 ; Waldenström and Vahlquist 1939). The homogeneity of such material has been questioned (Grinstein, Schwartz, and Watson 1945 ; Watson, Schwartz, and Hawkinson 1945 ; Prunty 1946). More recently (Nicholas and Rimington unpublished data) it was found on decarboxylation to yield a mixture of coproporphyrins I and III.

(*32*) *Phylloerythrin monomethyl ester.* Prepared by Professor C. Rimington from bile, m.p. 267 °C.

(*34*) *and* (*35*) *Methylphaeophorbide-a and -b.* From a mixture of phaeophorbins-*a* and -*b* (from chlorophyll after the removal of Mg) crystalline phaeophorbins-*a* and -*b* were prepared after the manner of Willstätter and Stoll (1928). The phaeophorbin-*a* was esterified by cold HCl-methanol, and the methylphaeophorbide-*a* was crystallized from ether and recrystallized from acetone-methanol, m.p. 227 °C. Methylphaeophorbide-*b* was similarly prepared from the crystalline phaeophorbin-*b*, and crystallized from acetone-methanol, m.p. 270 °C.

Compounds 14, 15, 28, 30, 31, and 32 were supplied by Professor C. Rimington.

III. SPECTROSCOPIC TECHNIQUE

A Perkin-Elmer Model 12B Spectrometer was used, with rock-salt optics for the study of the solid spectra and fluorite for accurate work on solutions in the 1600–1800 cm.$^{-1}$ and 3000–3600 cm.$^{-1}$ regions. The later spectra were taken after the instrument had been converted to a Model 12C by the installation of an A.C. thermocouple, A.C. amplifier, and chopped light beam.

Attempts to vacuum-sublime porphyrins on to a rock-salt plate were unsuccessful, as the specimens charred and decomposed. The substances were therefore studied as mulls in "Nujol". Since "Nujol" itself gives rise to strong absorption bands in the regions 1340–1400, 1430–1490, and 2700–3050 cm.$^{-1}$, the spectra in these regions have been omitted from the spectrograms, which are reproduced in Figure 1.

For spectra of solutions in the 1600–1800 cm.$^{-1}$ region a rock-salt cell about 0·5 mm. in thickness was used, and A.R. chloroform employed as solvent. In the 3000 cm.$^{-1}$ region quartz absorption cells 1–4 cm. in length were used, the solvent being A.R. carbon tetrachloride. Blank runs were carried out on the solvents alone.

IV. RESULTS

Figure 1 shows spectra for the solids over the range 4000–670 cm.$^{-1}$. Two substances (Nos. 17 and 35) were not available in sufficient quantities to enable a satisfactory spectrum of the solid to be obtained, but these were studied in chloroform solution in the 1600–1800 cm.$^{-1}$ region. Table 3 shows the frequencies in this region for all the compounds containing one or more carbonyl groups in addition to the propionic ester groups found in most of the porphyrins and haems. The esters only were measured in chloroform solution, since the free acids were not sufficiently soluble.

The frequencies measured in the 1600–1800 cm.$^{-1}$ region with the fluorite prism are more accurate (estimated error ± 2 cm.$^{-1}$) than those measured with the rock-salt prism (± 5 cm.$^{-1}$).

V. DISCUSSION

(a) *The Nature of the N–H Bonds in Porphyrins*

The intensity of the band corresponding to the N–H stretching frequency in porphyrins is weak and could only be detected when a sufficiently thick specimen was used. It occurs at 3280–3300 cm.$^{-1}$ with solid specimens.

Vestling and Downing (1939) found that aetioporphyrin (No. 2) in dilute carbon tetrachloride solution gave a small band at 3320 cm.$^{-1}$, which they attributed to the N–H stretching frequency, since it did not occur in the haem. In the present work solutions in carbon tetrachloride of protoporphyrin (No. 7), tetramethylhaematoporphyrin (No. 13), and the corresponding haems were studied. Only the porphyrins showed the band at 3320 cm.$^{-1}$. Now pyrrole, indole, and carbazole were all found to show their N–H vibration at 3500 cm.$^{-1}$ in dilute solution, so that a fair amount of N–H . . . N bonding within the porphyrin molecule is indicated. It is interesting to note that the N–H vibration frequency in pyrrole, indole, and carbazole falls by 100 cm.$^{-1}$ on passage from dilute solution to the solid or liquid state, whereas in the porphyrins it only falls by 20–40 cm.$^{-1}$. This indicates that the environment of the N–H bonds in the solid porphyrin is not greatly different from that in the isolated molecule,

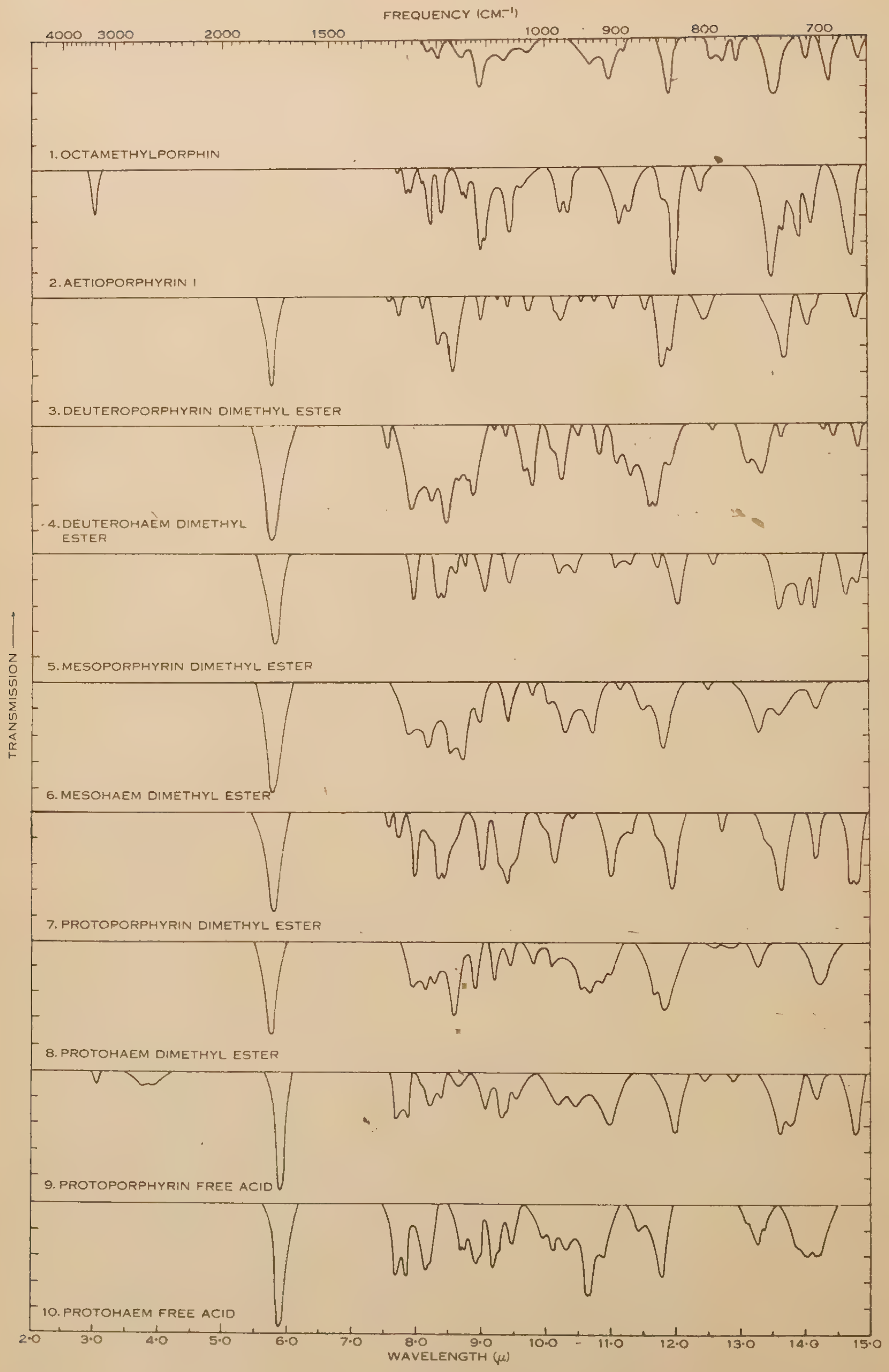

FREQUENCY (CM.⁻¹)
4000
3000
2000
1500
1000
900
800
700
1. OCTAMETHYLPORPHIN
2. AETIOPORPHYRIN I
3. DEUTEROPORPHYRIN DIMETHYL ESTER
4. DEUTEROHAEM DIMETHYL ESTER
5. MESOPORPHYRIN DIMETHYL ESTER
6. MESOHAEM DIMETHYL ESTER
7. PROTOPORPHYRIN DIMETHYL ESTER
8. PROTOHAEM DIMETHYL ESTER
9. PROTOPORPHYRIN FREE ACID
10. PROTOHAEM FREE ACID
TRANSMISSION
2·0
3·0
4·0
5·0
6·0
7·0
8·0
9·0
10·0
11·0
12·0
13·0
14·0
15·0
WAVELENGTH (μ)

Fig. 1

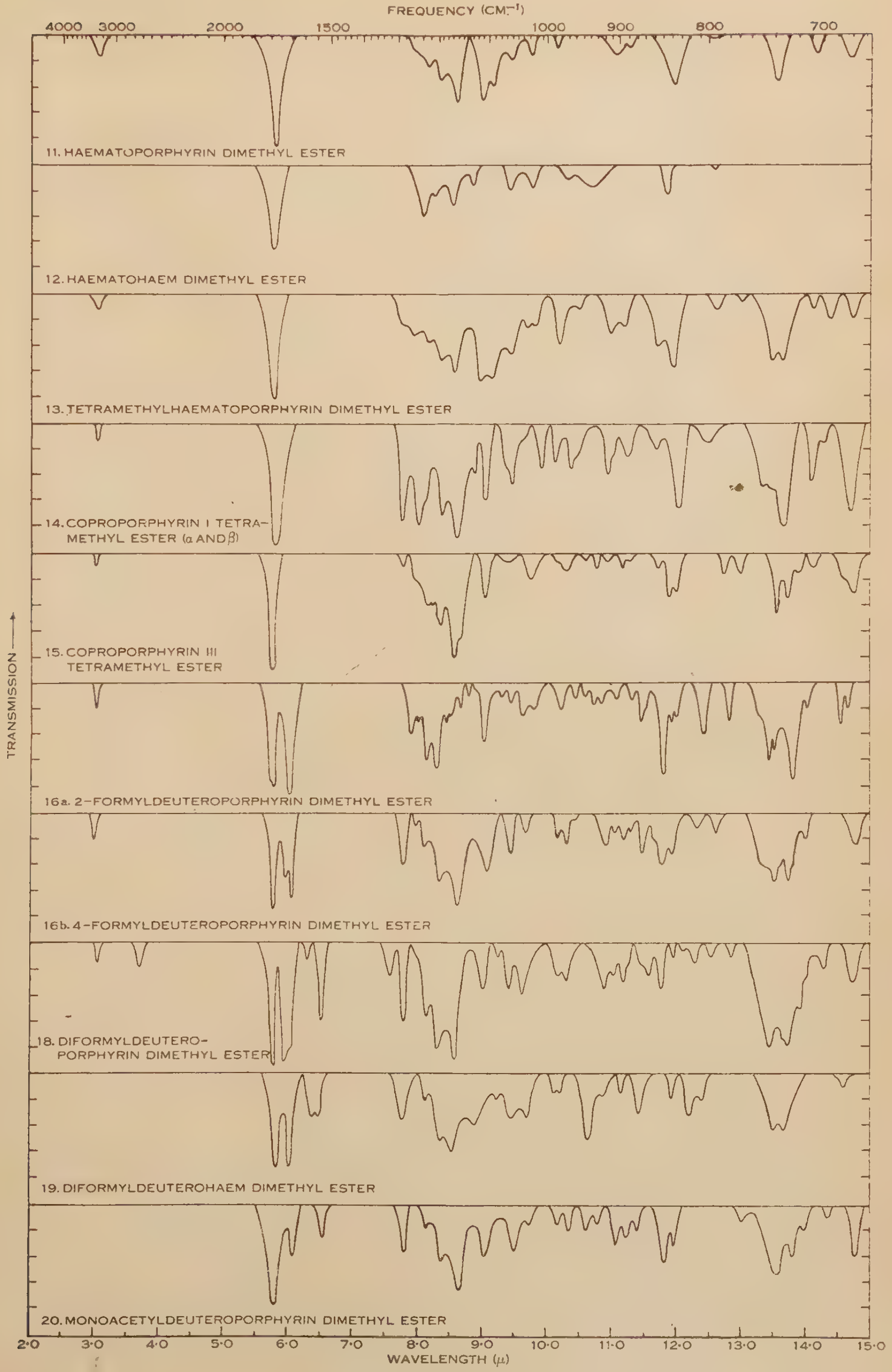


Fig. 1 (*Continued*)

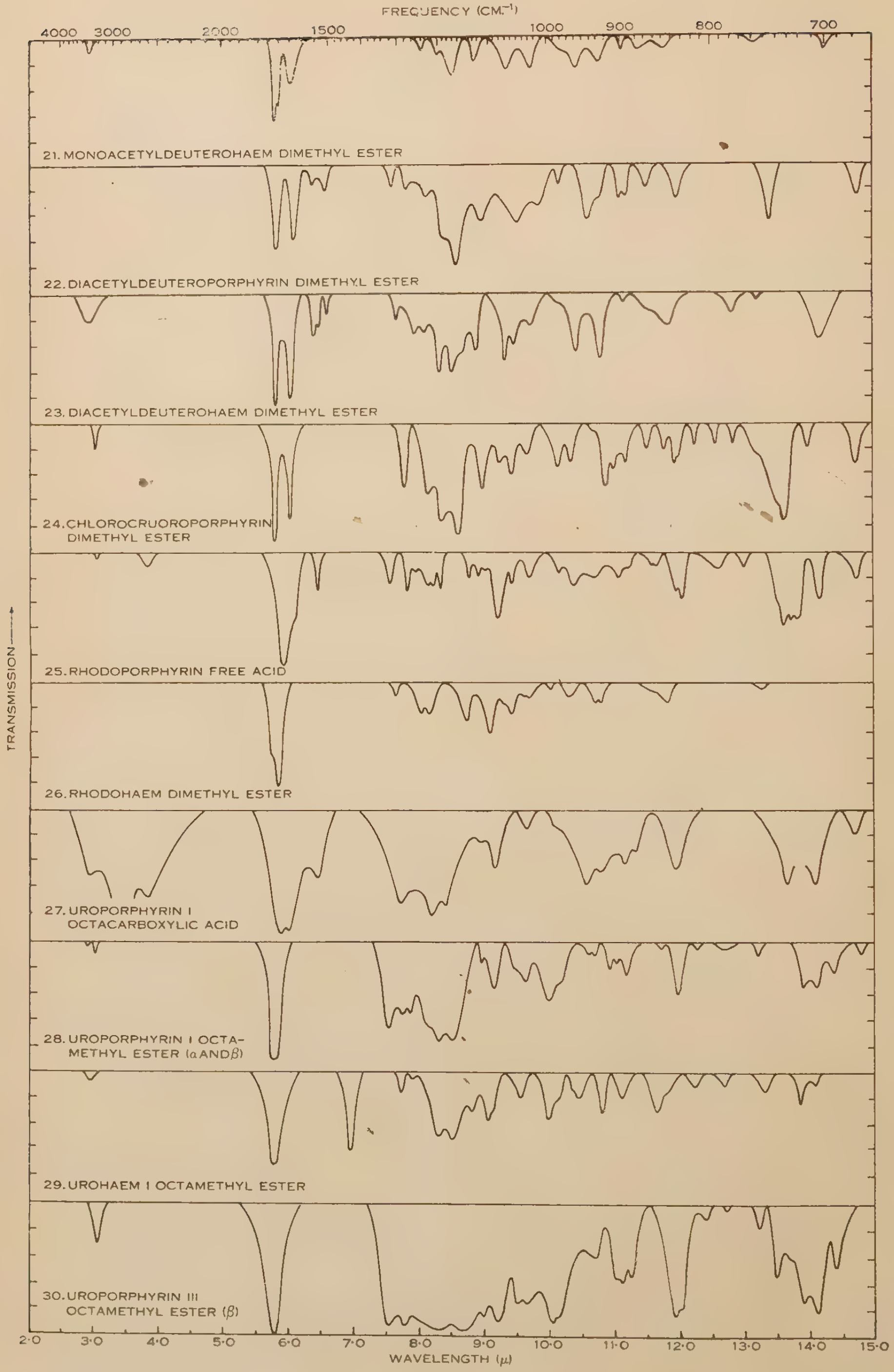


Fig. 1 (*Continued*)

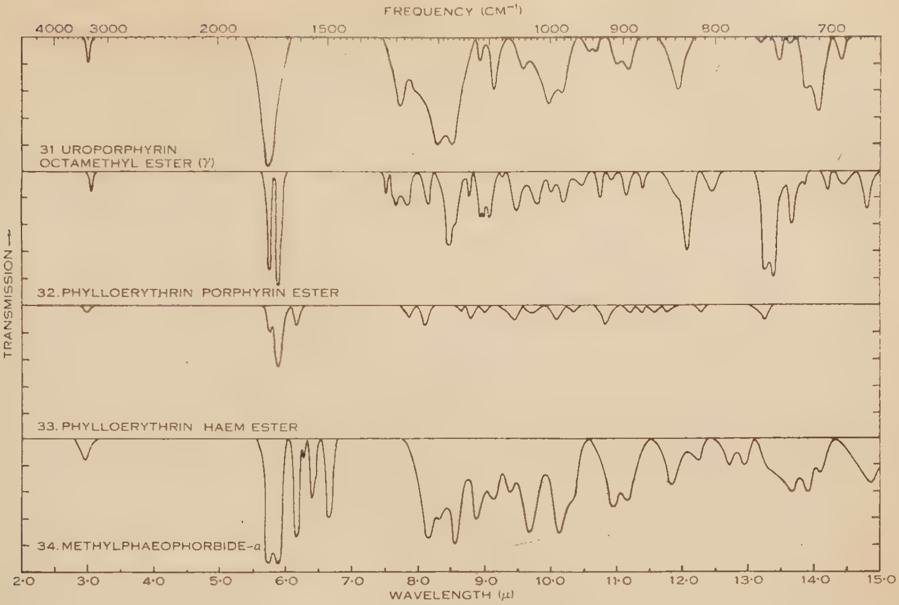


Fig. 1 (*Continued*)

and is confirmatory evidence that the N–H groups are occupied in intramolecular hydrogen bonding.

The above argument is based on the assumption that the vibration frequency of the N–H bond in porphyrins, if it were free, would be about the same as it is in pyrrole, and not unduly influenced by the nitrogen being part of an extensive conjugated system. The following considerations lend some weight to this assumption :

(i) The N–H vibration frequency in pyrrole remains unaltered by the addition of fused phenyl rings to the molecule to give indole and carbazole, in which a longer conjugated path is available to the electrons.

(ii) Theoretical investigations by Mulliken (1939) showed that pyrrole is not a fully " aromatic " molecule and that the " lone pair " of electrons on the nitrogen atom may not be fully conjugated. Simpson (1949), from similar arguments, has concluded that the porphyrins also are incompletely aromatic.

The =N–H in-plane bending frequency is less easy to locate than is the N–H stretching frequency. Lord and Miller (1942), in a complete vibrational analysis, assigned it a value of 1146 cm.$^{-1}$ in pyrrole, which is close to the value for hydrazoic acid (Davies 1939). In dimethylamine, however, it is assigned a value of 1500–1600 cm.$^{-1}$ (Walsh, unpublished data). In the porphyrin spectra there does not seem to be any band at a fairly constant frequency in either of these regions attributable to N–H bending. It may be observed, however, that in view of the evidence cited above for intramolecular hydrogen bonding, the N–H bending frequency would be expected to be higher in the porphyrins than in pyrrole. Shifts of 40–80 cm.$^{-1}$ towards higher frequencies were found

TABLE 3

SPECTRA IN THE 1600–1800 CM.$^{-1}$ REGION

No.	Compound	Absorption Frequencies of Solid : Rock-Salt Prism (cm.$^{-1}$)							Absorption Frequencies in Chloroform Solution : Fluorite Prism (cm.$^{-1}$)						
5	Mesoporphyrin							1730							1733
6	Mesohaem							1730							1734
28	Uroporphyrin I							1734							1735
29	Urohaem I							1734							1735
16*a*	2-Monoformyldeuteroporphyrin					1655		1734				1658			1734
16*b*	4-Monoformyldeuteroporphyrin					1654		1736				1659			1734
17	Monoformyldeuterohaem					1665		1732				1669			1735
18	Diformyldeuteroporphyrin	1533		1584*	1658†	1673		1736				1657	1674†		1734
19	Diformyldeuterohaem	1545		1560		1663		1726				1659			1734
20	Monoacetyldeuteroporphyrin	1542				1644		1735			‡				
21	Monoacetyldeuterohaem					1656	1718†	1736				1665			1734
22	Diacetyldeuteroporphyrin	1520		1571		1650		1736				1658			1735
23	Diacetyldeuterohaem	1543		1558		1656		1732				1662			1735
24	Chlorocruoroporphyrin					1654		1734				1658			1733
32	Phylloerythrin porphyrin						1690	1735					1691		1732
33	Phylloerythrin haem						1696	1732		1620	1636		1694		1735*
34	Methylphaeophorbide-*a*	1543	1556	1584	1620		1698	1736	1584	1619			1699		
35	Methylphaeophorbide-*b*					1654	1700	1736	1589	1604		1662	1702	1719	1735

* Weak band. † Shoulder. ‡ Specimen not available for study with the fluorite prism.

by Davies (1948) for the O–H bending vibration in phenol and methanol on transition from dilute to concentrated solutions, and were ascribed by him to increased hydrogen bonding in the liquid state.

In the spectra of some of the porphyrins and haems weak bands are found at 3370–3440 cm.$^{-1}$. These can be interpreted as combination or overtone bands (usually involving the C=O vibration) and are not to be confused with N–H vibrations.

(*b*) *The Distinction of Position Isomers*

The spectra shown for coproporphyrin I and III (Nos. 14 and 15) agree well with those published by Gray, Neuberger, and Sneath (1950). There are distinct differences between the spectra of the two isomers in the range 700–1400 cm.$^{-1}$, particularly near 750, 1060, 1250, and 1290 cm.$^{-1}$. It is hoped to make use of these differences to develop a method for the determination of the I and III isomers in the presence of each other. Such a method would be particularly useful in the examination of the porphyrins excreted by patients suffering from porphyrias of the various types. The isolation from urine or faeces of small amounts of coproporphyrins I and III, free of other substances, is not very difficult. Separation of the isomers on a scale too small for the employment of fractional crystallization is, however, very difficult, and no means of doing this quantitatively is yet available, nor is there any satisfactory alternative method for determining the proportions of isomers in a mixture. Separation of the esters by chromatography on alumina, with elution of the type III isomer by 30 per cent. aqueous acetone, was claimed by Watson and Schwartz (1940), but Jope and O'Brien (1945) and Helwig (1949) were unable to confirm this. Schwartz, Hawkinson, and Watson (1946) and Schwartz *et al.* (1947) reported another method, depending on the much more rapid quenching of the fluorescence, in 30–35 per cent. aqueous acetone at reduced temperature, of the I isomer than of the III isomer. Other workers (Weatherall, Prunty, personal communications) have found this method difficult to apply.

Samples α and β of uroporphyrin I (No. 28) gave identical spectra which, however, differ slightly from that of the IIIβ isomer (No. 30), particularly in the low-frequency region, though the difference is not so great as between the coproporphyrin I and III isomers. Uroporphyrin sample γ (No. 31) is a typical " Waldenström ester ". The evidence for the inhomogeneity of such material, summarized above, is confirmed by comparison of its spectrum with those of uroporphyrin Iα and of IIIβ. In the region of 1200 cm.$^{-1}$ the spectrum resembles uroporphyrin I rather than III, while in the region of 670–750 cm.$^{-1}$ its spectrum is similar to that of uroporphyrin III. The spectrum of sample γ can be interpreted as a composite of the spectra of Iα and IIIβ.

In spite of the fact that insufficient quantities of urohaem III were available to give a good spectrum, differences between it and the haem I were noted in the low-frequency region.

(*c*) *Characteristic Frequencies of the Carbonyl Group*

It is seen from Table 3 that all porphyrins and haems containing a propionic ester group show a C=O frequency which is very constant at 1730–1735 cm.$^{-1}$, and is not appreciably affected by transition from solution to the solid state. Since the ester carbonyl group is well " insulated " from the rest of the molecule by two methylene groups, the frequency is independent of the structure of the rest of the molecule.

Only one specimen containing an ester group directly attached to a heterocyclic ring has yet been studied, viz. the diester of rhodohaemin (No. 26). The shoulder at about 1708 cm.$^{-1}$ on the low-frequency side of the propionic ester band is probably due to the carbonyl stretching frequency of the COOMe group on the pyrrole ring. The lower value of this frequency, as compared with that for a side-chain ester, is due to the conjugation of the ester C=O group with the double bonds of the heterocyclic ring (cf. Hartwell, Richards, and Thompson 1948).

Phylloerythrin (porphyrin and haem) esters (Nos. 32 and 33) and methylphaeophorbide-*a* and -*b* (Nos. 34 and 35) show a band at 1690–1700 cm.$^{-1}$, which is evidently due to the keto group present in the *cyclo*pentanone ring in these compounds. The ester group on this ring evidently has the same carbonyl frequency as the side-chain ester, viz. 1730–1735 cm.$^{-1}$.

The free carboxylic acid group has a frequency considerably lower than that of an ester, and can be identified at 1683–1700 cm.$^{-1}$. Here again, conjugation with the pyrrole ring leads to a lower frequency than would be expected for an aliphatic acid.

It will be seen that formyl and acetyl groups substituted in the pyrrole nucleus give rise to absorption at 1658–1665 cm.$^{-1}$, accompanied by a band at 1560–1580 cm.$^{-1}$ if there are two such groups. Methylphaeophorbide-*a* (No. 34) also shows a band at 1584 cm.$^{-1}$. Hartwell, Richards, and Thompson (1948) found that the C=O frequency was usually higher in aldehydes than in ketones, but the compounds examined here show little evidence of this. There seems to be a tendency, however, for the carbonyl vibration in the haem to be at a slightly higher frequency than in the porphyrin.

Fortunately, aldehyde side-chains on the porphyrin nucleus can be distinguished from acetyl side-chains and from keto groups in isocyclic rings through the difference in reactivity of the respective haemochromagens towards hydroxylamine or hydrazine, as described by Lemberg and Falk (1951).

(*d*) *Absorption due to the Vinyl Group*

The following compounds containing the vinyl side-chain on the pyrrole nucleus were studied : protoporphyrin and haem (ester and free acid Nos. 7–10), chlorocruoroporphyrin, and methylphaeophorbide-*a* and -*b* (Nos. 34 and 35).

Double bonds in aliphatic structures give moderate absorption at about 1630–1650 cm.$^{-1}$, the frequency being lowered somewhat by conjugation (Blout, Fields, and Karplus 1948). No absorption in this region was found for protoporphyrin or haem. The bands at 1619 cm.$^{-1}$ in methylphaeophorbide-*a* and -*b* may be due to the C=C vibration, although if this is so it is difficult to see why there should be no C=C absorption in protoporphyrin and haem.

(*e*) *The Low-Frequency Region*

Below about 1300 cm.$^{-1}$ it is difficult to assign the absorption bands to specific vibrations, as these vibrations involve complicated motions of more than one pair of atoms. It does not seem profitable at this stage to try to draw empirical correlations from the spectra in this region of such large molecules as porphyrins and haems, as there is little infra-red spectral data on simple pyrrole derivatives to form a basis for comparison. It is not even easy to see any systematic differences between porphyrins and haems, though it may be noted that the 1105–1110 cm.$^{-1}$ band which occurs in almost all porphyrins is shifted to about 1130 cm.$^{-1}$ in the haems.

As mentioned above, the =N—H in-plane bending frequency should be found in this region, but there seems no indication of any strong band which is systematically present in the porphyrins and absent in the haems.

VI. Acknowledgments

The authors are greatly indebted to Dr. R. Lemberg and to Professor C. Rimington for much helpful discussion during the progress of this investigation, for providing facilities and accommodation, and for supplying some of the samples. Professor S. Goldschmidt and Dr. J. H. Hale are also thanked for making materials available. One of the authors (J.E.F.) was supported during this work by a Fellowship from the National Health and Medical Research Council of Australia, and later in London, by a Fellowship from the Nuffield Foundation. The Beckman spectrophotometer was purchased out of a grant to Professor Rimington from the Central Research Funds Committee of the University of London.

VII. References

Blout, E. K., Fields, M., and Karplus, R. (1948).—*J. Amer. Chem. Soc.* **70** : 194.
Davies, M. M. (1939).—*Trans. Faraday Soc.* **35** : 1184.
Davies, M. M. (1948).—*J. Chem. Phys.* **16** : 267, 274.
Fischer, H., and Röse, H. (1913).—*Ber. dtsch. chem. Ges.* **46** : 2460.
Gray, C. H., and Holt, L. B. (1948).—*Biochem. J.* **43** : 191.
Gray, C. H., Neuberger, A., and Sneath, P. H. A. (1950).—*Biochem. J.* **47** : 87.
Grinstein, M. (1947).—*J. Biol. Chem.* **167** : 515.
Grinstein, M., Schwartz, S., and Watson, C. J. (1942).—*J. Biol. Chem.* **157** : 323.
Hartwell, E. J., Richards, R. E., and Thompson, H. W. (1948).—*J. Chem. Soc.* **1948** : 1436.
Helwig, F. (1949).—*Z. Ges. Inn. Med.* **4** : 415.
Jope, E. M., and O'Brien, J. R. P. (1945).—*Biochem. J.* **39** : 239.
Lemberg, R., and Falk, J. E. (1951).—*Biochem. J.* **49** : 674.
Lemberg, R., Falk, J. E., Rawlinson, W. A., Hale, J. H., and Rimington, C. (1949).—Abstr., 1st Int. Congr. Biochem.
Lord, R. C., and Miller, F. A. (1942).—*J. Chem. Phys.* **10** : 328.
Mason, F. A. (1925).—*Chem. and Ind.* **44** : 577.
Mulliken, R. S. (1939).—*J. Chem. Phys.* **7** : 339.
Nicholas, R. E. H., and Rimington, C. (1951).—*Biochem. J.* (in press).
Paul, K. G. (1951).—*Acta Chem. Scand.* **5** : 328.
Prunty, F. T. G. (1946).—*Arch. Intern. Med.* **77** : 623.
Rawlinson, W. A., and Hale, J. H. (1949).—*Biochem. J.* **45** : 247.
Rimington, C. (1936).—*Onderstepoort J. Vet. Sci.* **7** : 567.

Rimington, C. (1939).—*Proc. Roy. Soc.* B **127** : 106.

Rimington, C., and Miles, P. A. (1951).—*Biochem. J.* (in press).

Schumm, O. (1928).—*Hoppe-Seyl. Z.* **178** : 11.

Schwartz, S., Hawkinson, V., Cohen, S., and Watson, C. J. (1947).—*J. Biol. Chem.* **168** : 133.

Schwartz, S., Hawkinson, V., and Watson, C. J. (1946).—*Science* **103** : 338.

Simpson, W. T. (1949).—*J. Chem. Phys.* **17** : 1218.

Stern, A., and Wenderlein, H. (1934).—*Z. Phys. Chem.* A **170** : 345.

Stern, A., and Wenderlein, H. (1935*a*).—*Z. Phys. Chem.* A **174** : 81.

Stern, A., and Wenderlein, H. (1935*b*).—*Z. Phys. Chem.* A **174** : 321.

Stern, A., and Wenderlein, H. (1935*c*).—*Z. Phys. Chem.* A **175** : 405.

Vestling, C. S., and Downing, J. R. (1939).—*J. Amer. Chem. Soc.* **61** : 3511.

Waldenström, J. (1937).—*Acta. Med. Scand.* **72** (Suppl.).

Waldenström, J., and Vahlquist, B. (1939).—*Hoppe-Seyl. Z.* **260** : 189.

Warburg, O., and Negelein, E. (1932).—*Biochem. Z.* **244** : 9.

Watson, C. J., and Schwartz, S. (1940).—*Proc. Soc. Exp. Biol. Med. N.Y.* **44** : 7.

Watson, C. J., Schwartz, S., and Hawkinson, V. (1942).—*J. Biol. Chem.* **157** : 345.

Willstätter, R., and Stoll, A. (1928).—" Investigations on Chlorophyll." Trans. Schertz, F. M., and Merz, A. R. (Science Press Printing Co. : Lancaster, Pa.)

POLAROGRAPHY WITH ALTERNATING CURRENTS

III. A.C. POLAROGRAPHY OF COPPER, BISMUTH, ANTIMONY, AND TIN WITH SPECIAL REFERENCE TO THE PRESENCE OF HALOGEN IONS

By B. Breyer,* F. Gutmann,† and S. Hacobian*

[*Manuscript received April* 13, 1951]

Summary

The A.C. polarography of copper, bismuth, antimony, and tin is discussed with special reference to the effect of the presence of halogen ions on the reversibility of the electrode reaction. While Cu^{++}, Bi^{+++}, and Sb^{+++} ions are irreversibly reduced in perchloric acid as supporting electrolyte, the electrode process becomes reversible when small amounts of halogen ions are added. It is shown that this cannot be explained by assuming the formation of complex halogeno metal ions. Sn^{++} ions in perchloric acid are reversibly reduced, but only in the presence of a large excess of chloride ions, indicating a reduction of the stannous tetrachloro complex ion ($SnCl_4^{=}$), as already observed by previous workers.

A new polarographic cell design suitable for A.C. polarography (the " pool-condenser " saturated calomel electrode) is described in Appendix I.

I. Introduction

The theory and practice of A.C. polarography has been discussed in two previous publications (Breyer, Gutmann, and Hacobian 1950*a*, 1950*b*). The present paper deals with the A.C. polarographic estimation of copper, bismuth, antimony, and tin, and shows that the choice of a suitable supporting electrolyte is essential in order to obtain satisfactory results. Thus, for example, these ions were reduced reversibly in 0·1N HCl or KCl as supporting electrolyte, and good A.C. waves were obtained, but in perchloric acid free of halogen ion the reduction process became irreversible as shown by the D.C. " log plots " and by the lowering of A.C. waves. These experiments showed that the presence of halogen ions is needed for the reversibility of the electrode process, a fact already observed by Heyrovsky (1947).

II. Copper

In 0·5N perchloric acid as supporting electrolyte copper ions gave practically no A.C. wave. D.C. polarography showed an irreversible step at 0 V. *v.* the saturated calomel electrode (S.C.E.) (cf. Table 1). In 0·5N sodium perchlorate, however, a nearly reversible D.C. step was obtained as shown by the linear log plot of 33 mV. slope. A corresponding A.C. wave at the same potential of about 0 V. *v.* S.C.E. was recorded. Table 2 shows that the nature of the

---

*Physico-Chemical Laboratory, Department of Agriculture, University of Sydney.

† Department of Electrical Engineering, New South Wales University of Technology, Sydney.

TABLE 1

THE INFLUENCE OF HALOGEN IONS ON THE REVERSIBILITY OF THE Cu^{++} POLAROGRAPHIC WAVE (IN ABSENCE OF AIR)

A.C. 15 mV. r.m.s. ; T, 15 °C.

Cu^{++} Concn. (N)	Supporting Electrolyte	Halogen Ions Added	Halogen Ion Concn. (N)	$i_D\sim$* (μA. r.m.s.)	i_d (μA.)	E_s *v.* S.C.E. (V.)	$E_{\frac{1}{2}}$ *v.* S.C.E. (V.)	Log Plot of D.C. Pol. (mV. slope)†
10^{-3}	0·5N $HClO_4$	—	—	<1·0	3·0	—	—0·005	Non-linear
,,	,,	Cl^- (HCl)	5×10^{-4}	1·5	3·0	—0·005	0·00	Non-linear
,,	,,	,,	2×10^{-3}	7·9	2·88	—0·015	+0·01	38†
,,	,,	,,	10^{-2}	12·0	3·0	—0·065	—	—‡
$1{\cdot}3\times10^{-3}$	,,	Br^- (KBr)	10^{-5}	1·5	3·65	+0·03	+0·015	Non-linear
,,	,,	,,	10^{-4}	8·2	3·45	+0·02	+0·02	38†
10^{-3}	,,	I^- (KI)	10^{-5}	2·4	2·88	0·00	0·00	Non-linear
,,	,,	,,	2×10^{-5}	3·0	2·84	+0·005	+0·005	45†
,,	,,	,,	2×10^{-4}	8·0	2·82	+0·025	0·00	40†
,,	,,	,,	6×10^{-4}	11·2	—	—0·08	—	—‡

* Reference to symbols used : $i_D\sim$, maximum differential current (A.C.) ; i_d, D.C. diffusion current ; E_s, summit potential of A.C. wave ; $E_{\frac{1}{2}}$, D.C. half-step potential.

† Theoretical slope : 29 mV.

‡ Halogen ion step overlapping.

TABLE 2

A.C. POLAROGRAPHY OF 10^{-3}N Cu^{++} IN VARIOUS SUPPORTING ELECTROLYTES (IN AIR) SHOWING THE DIFFERENT DEGREES OF REVERSIBILITY OF THE REDUCTION PROCESS

A.C. 15 mV. r.m.s. ; T, 20 °C.

Supporting Electrolyte	$i_D\sim$ (μA. r.m.s.)	Log Plot of D.C. Pol. (mV. slope)*
0·5N Na_2SO_4	14·5	31
0·5N K_2SO_4	14·0	31
0·5N NH_4ClO_4	10·5	33
0·5N $NaClO_4$	10·0	33
0·5N CH_3COONa	12·0	34
0·5N KNO_3	11·5	36
0·5N $HClO_4$	<1·0	Non-linear
0·5N H_2SO_4	13·5	32
0·5N HNO_3	7·5	40

* Theoretical slope : 29 mV.

cation as well as the anion present in the supporting electrolyte affects the reversibility of the reaction

$$Cu^{++}+2e \rightarrow Cu.$$

This is indicated by the magnitudes of the maximum differential currents and the D.C. log plots, obtained by plotting log $i_d - i/i$, i.e. log [Ox]/[Red], against the applied D.C. potential, E.

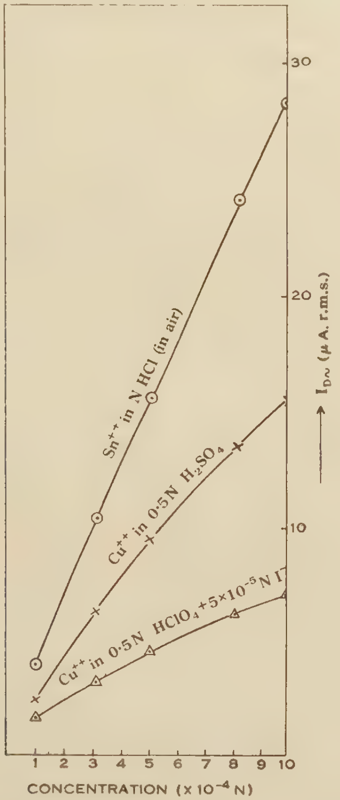


Fig. 1.—Calibration curves of Cu^{++} and Sn^{++} in different supporting electrolytes and in the presence of air. A.C. 15 mV. r.m.s.; T, 20 °C.

Table 2 shows clearly the importance of the reversibility of the electrode process for the production of good A.C. waves. In the presence of perchloric acid and halogen ions, however, copper ions were reversibly reduced and good A.C. waves were consequently obtained. The reversibility of the electrode process (as determined by the magnitude of $i_{D\sim}$) increased with increased addition of halogen ions. Table 1 shows the relationship between $i_{D\sim}$ values of Cu^{++} and the amounts of halogen ions added.

That the Cu^{++} ions are not reduced as halogeno complex is proved by the following experimental evidence:

(i) While 10^{-3}N Cu^{++} was irreversibly reduced in 0·5N perchloric acid, the process became nearly reversible in the presence of 2×10^{-5}N iodide ion. This is shown by the presence of an A.C. wave and a marked change in the D.C. log plot (cf. Table 1).

(ii) The half-step potential of Cu^{++} in 0·5N perchloric acid was shifted to more positive values after the addition of iodide ion.

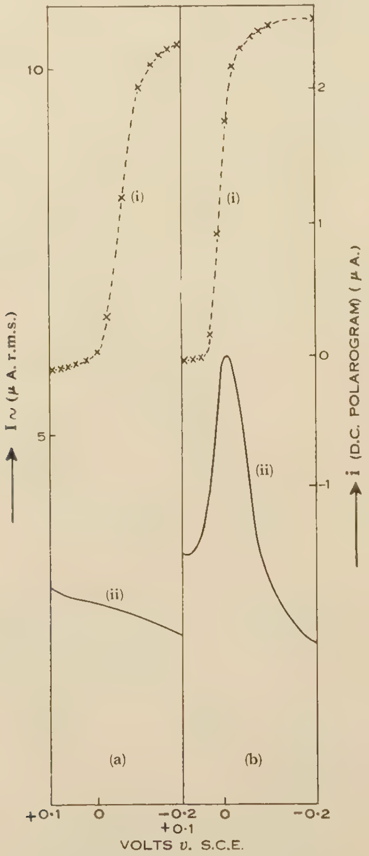


Fig. 2.—Influence of Cl^- on the Bi^{+++} wave.
(*a*) (i) D.C. polarogram of 10^{-3}N Bi^{+++} in 0·5N $HClO_4$ as supporting electrolyte; T, 20 °C.
(ii) A.C. polarogram of 10^{-3}N Bi^{+++} in 0·5N $HClO_4$. A.C. 15 mV. r.m.s.; T, 20 °C.
(*b*) (i) As in 2 (*a*) (i) plus 10^{-3}N Cl^-.
(ii) As in 2 (*a*) (ii) plus 10^{-3}N Cl^-.

(iii) Varying degrees of reversibility of the copper ion reduction process were also observed in supporting electrolytes other than perchloric acid (Table 2).

Experiment (i) proves that the reduction of the copper ion cannot proceed via the iodo complex: the halogen ion concentration is much too small to account for a complex formation to any appreciable extent. Experiment (ii) indicates that the reduction of copper as complex ion is unlikely: with rare exceptions the reduction of complex ions entails a negative shift of the reduction potential. Experiment (iii) shows that the varying degree of reversibility of the electrode process is determined not only by the presence of halogen ions alone but also by the nature of the cations in the supporting electrolyte.

Copper can be successfully estimated by A.C. polarography in suitable supporting electrolytes and also in the presence of air (cf. Table 2). The maximum differential current values for copper ions are approximately linearly proportional to the ion concentration over a range of 10^{-3}–10^{-4}N (cf. Fig. 1).

III. BISMUTH AND ANTIMONY

In 0·5N perchloric acid both Bi^{+++} and Sb^{+++} ions (actually BiO^+ and SbO^+) were irreversibly reduced. This is indicated by the D.C. log plot and by the virtual absence of A.C. waves (cf. Tables 3 and 4).

TABLE 3

INFLUENCE OF HALOGEN IONS ON THE REVERSIBILITY OF THE POLAROGRAPHIC WAVES OF Bi^{+++}

A.C. 15 mV. r.m.s.; T, 15 °C.

Bi^{+++} Concn. (N)	Supporting Electrolyte	Halogen Ions Added	Halogen Ion Concn. (N)	$i_{D\sim}$ (μA. r.m.s.)	i_d (μA.)	E_s v. S.C.E. (V.)	$E_{\frac{1}{2}}$ v. S.C.E. (V.)	Log Plot of D.C. Pol. (mV. slope)†
10^{-3}	0·5N $HClO_4$	—	—	—	2·30	—	—0·045	52
,,	,,	Cl^- (HCl)	2×10^{-3}	4·3	2·31	+0·003	0·00	22·4
,,	,,	,,	10^{-2}	24·0	2·25	—0·007	—0·01	22·0
,,	,,	,,	10^{-1}	55	2·25	—0·056	—0·045	22·0
,,	,,	Br^- (KBr)	5×10^{-5}	14	2·26	+0·017	+0·015	22·0
,,	,,	,,	5×10^{-4}	25	2·50	+0·010	+0·015	21·0
,,	,,	I^- (KI)	5×10^{-5}	10·5	2·30	+0·01	+0·015	22·0
,,	,,	,,	5×10^{-4}	28	(1·80)*	+0·017	+0·015	—*

* D.C. polarogram not well defined; the i_d values here are very approximate.

† Theoretical slope: 19 mV.

Bismuth and antimony* ions, like those of copper, gave reversible A.C. polarographic waves in 0·5N $HClO_4$ when in the presence of halogen ions.

Tables 3 and 4 show how the reversibility of the electrode process depends upon the concentrations of halogen ions added. Here again it can be seen that the magnitude of $i_{D\sim}$ depends on the reversibility of the electrode process. These tables indicate that 10^{-3}N Bi^{+++} in the presence of 5×10^{-5}N Br^- and 5×10^{-5}N I^-, for example, gave quite significant A.C. waves ($i_{D\sim}$, 14·0 and 10·5 μA. r.m.s. respectively). Exactly as with copper and for the same reasons, electrodeposition of bismuth from the halogen complex can be excluded.

---

* The antimony solution contained a mixture of 0·5N $HClO_4$ and 0·1N $NaClO_4$ as supporting electrolyte. The latter was introduced during the preparation of the antimony perchlorate from Sb_2O_3, dissolving in NaOH and acidifying with $HClO_4$.

The appearance of an A.C. wave on the addition of hydrochloric acid (Figs. 2 and 3, Tables 3 and 4) shows that both the bismuth and antimony ion reductions are thereby made reversible. Sb^{+++} in perchloric acid-sodium perchlorate mixture, free of halogen ion, gave two D.C. polarographic steps separated by about 150 mV. (cf. Fig. 3 (*a*)). By adding chloride ions the two steps gradually merged into one with a theoretical slope corresponding to a three electron reduction process.

TABLE 4

INFLUENCE OF HALOGEN IONS ON THE REVERSIBILITY OF THE POLAROGRAPHIC WAVES OF Sb^{+++}
A.C. 15 mV. r.m.s.; T, 15 °C.

Sb^{+++} Concn. (N)	Supporting Electrolyte	Halogen Ions Added	Halogen Ion Concn. (N)	$i_{D\sim}$ (μA. r.m.s.)	i_d (μA.)	E_s v. S.C.E. (V.)	$E_{\frac{1}{2}}$ v. S.C.E. (V.)	Log Plot of D.C. Pol. (mV. slope)*
10^{-3}	$0\cdot5N\ HClO_4$ + $0\cdot1N\ NaClO_4$	—	—	—	1·00 2·05†	—	—0·175 —0·275†	Non-linear
,,	,,	Cl^- (HCl)	10^{-3}	—	2·64	—	—‡	Non-linear
,,	,,	,,	10^{-1}	3·9	2·64	—0·15	—0·15	30
,,	,,	,,	1	36	2·62	—0·185	—0·17	20
,,	,,	Br^- (KBr)	2×10^{-3}	4·5	2·90	—0·140	—0·145	32
,,	,,	,,	10^{-2}	29	2·78	—0·147	—0·145	20
,,	,,	I^- (KI)	10^{-4}	33	3·50**	—0·145	—0·148	23
,,	,,	,,	4×10^{-4}	45	4·25**	—0·140	—0·142	20

* Theoretical slope, 19 mV.

† Two polarographic waves observed.

‡ Steps too close for measurement of individual $E_{\frac{1}{2}}$ values.

** These values appear higher than the others because of the overlapping of the anodic iodide wave with the cathodic Sb^{+++} step.

IV. TIN

Stannous ion in 0·5N perchloric acid gives no D.C. polarogram before the discharge of hydrogen (Kolthoff and Lingane 1941, p. 263). In the presence of a large excess of Cl^- ions, however, Sn^{++} produced a well-defined A.C. polarogram as shown in Figure 4. The summit potential was practically identical with the D.C. half-step potential. In contradistinction to copper, bismuth, and antimony, stannous ions required a comparatively large excess of chloride ions to obtain good A.C. waves. This is in accordance with the fact that the $(SnCl_4)^-$ ion is reduced reversibly (Smrz 1925).

The calibration curve for tin in hydrochloric acid as supporting electrolyte and in the presence of air is shown in Figure 1. A nearly linear relationship between $i_{D\sim}$ and ion concentration was obtained over a more than tenfold dilution range ($2\cdot10^{-3}$–10^{-4}N).

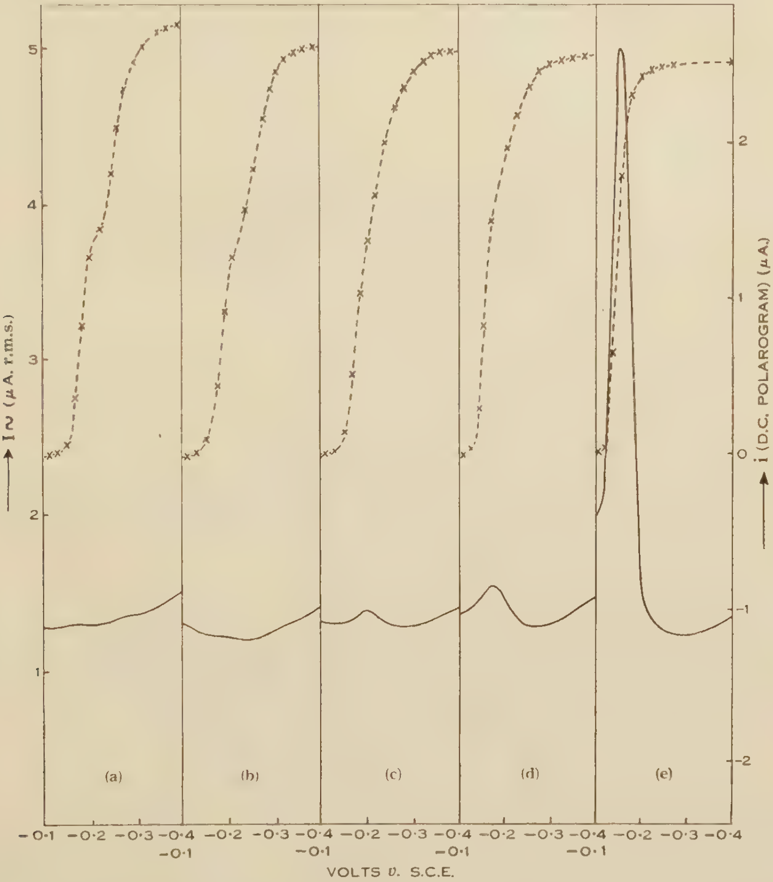


Fig. 3.—Influence of Cl^- on the A.C. and D.C. polarographic waves of Sb^{+++}. Supporting electrolyte, $0\cdot5N\ HClO_4 + 0\cdot1N\ NaClO_4$ (absence of air). A.C. 15 mV. r.m.s.; T, 20 °C.

(*a*) A.C. and D.C. polarograms of $10^{-3}N\ Bi^{+++}$.
(*b*) 3 (*a*) plus $10^{-3}N\ Cl^-$.
(*c*) 3 (*a*) plus $2\times10^{-3}N\ Cl^-$.
(*d*) 3 (*a*) plus $10^{-2}N\ Cl^-$.
(*e*) 3 (*a*) plus $10^{-1}N\ Cl^-$.

V. ACKNOWLEDGMENT

The authors wish to thank the Trustees of the Science and Industry Endowment Fund for financial assistance.

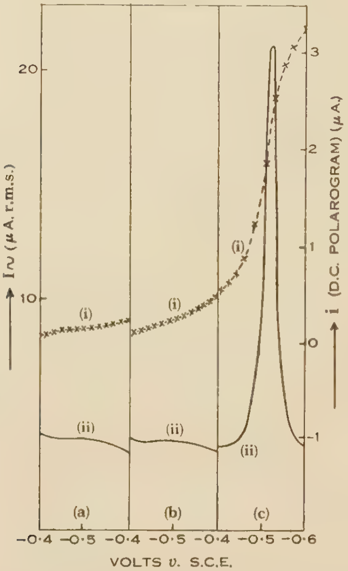


Fig. 4.—Polarographic waves of the stannous chloro complex ion. Supporting electrolyte, 0·5N $HClO_4$ (absence of air). A.C. 15 mV. r.m.s.; T, 20 °C.

(*a*) (i) D.C. polarogram of 10^{-3}N Sn^{++}.
(ii) A.C. polarogram of 10^{-3}N Sn^{++}.
(*b*) (i) D.C. polarogram of 10^{-3}N Sn^{++} plus 0·1N Cl^- (HCl).
(ii) A.C. polarogram of 10^{-3}N Sn^{++} plus 0·1N Cl^- (HCl).
(*c*) (i) D.C. polarogram of 10^{-3}N Sn^{++} plus 0·5N Cl^- (HCl).
(ii) A.C. polarogram of 10^{-3}N Sn^{++} plus 0·5N Cl^- (HCl).

VI. REFERENCES

BREYER, B., GUTMANN, F., and HACOBIAN, S. (1950*a*).—*Aust. J. Sci. Res.* A **3**: 558.

BREYER, B., GUTMANN, F., and HACOBIAN, S. (1950*b*).—*Aust. J. Sci. Res.* A **3**: 567.

HEYROVSKY, J. (1947).—*Faraday Soc. Disc.* No. 1: 219.

KOLTHOFF, I. M., and LINGANE, J. J. (1941).—"Polarography." 1st Ed. (Interscience Publishers Inc.: New York.)

SMRZ, J. (1925).—*Rec. Trav. Chim.* **44**: 580.

APPENDIX I

The "Pool-Condenser" Saturated Calomel Electrode

A new polarographic cell arrangement suitable for A.C. polarography and incorporating the saturated calomel electrode proved very useful. This cell arrangement has the advantage of having a relatively low A.C. impedance and it is proposed to call it for convenience the "pool-condenser" saturated calomel electrode ((pc) S.C.E.). The apparatus is shown in Figure 5. If the capacity of the condenser *C* is large (e.g. 50–100 μF.), the A.C. flows mainly through the condenser and by-passes the high impedance of the reference electrode and salt bridges whereas the D.C. flows through the salt bridges and the reference half cell. In all the experiments reported in this paper the (pc) S.C.E. with an electrolytic condenser of 100 μF. capacity was used. With a 50 c/s. A.C. signal input the impedance of the condenser is approximately of the order of 32 ohms, which is less than the impedance of the supporting electrolyte in the cell (about 90 ohms for 0·5N $HClO_4$). These impedance values are far smaller than the ohmic resistance of the saturated calomel electrode and the salt bridges which is of the order of thousands of ohms.

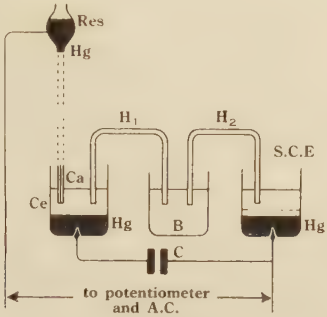


Fig. 5.—The pool-condenser saturated calomel electrode. *Hg*, mercury reservoir; *Ca*, dropping mercury electrode; *Ce*, polarographic cell; H_1, bridge containing supporting electrolyte; H_2, bridge containing saturated KCl; *B*, beaker containing saturated KCl; *S.C.E.*, saturated calomel half-cell; *C*, 100 μF. electrolytic condenser.

The summit potentials for various ion species determined *v.* S.C.E. were practically identical with those determined *v.* the (pc) S.C.E. as shown in Table 5.

TABLE 5
VALUES OF SUMMIT POTENTIALS *v.* S.C.E. AND *v.* (pc) S.C.E. FOR Bi^{+++}, Cd^{++}, Pb^{++}, AND Zn^{++} IN 0·5N HCl (ALL 10^{-3}N)

Ions	E_s *v.* S.C.E. (V.)	E_s *v.* (pc) S.C.E. (V.)
Bi^{+++}	—0·118	—0·116
Cd^{++}	—0·605	—0·608
Pb^{++}	—0·400	—0·400
Zn^{++}	—1·05	—1·05

POLAROGRAPHY WITH ALTERNATING CURRENTS

IV. A.C. POLAROGRAPHY OF URANIUM, IRON, NICKEL, AND MANGANESE

By B. Breyer* and S. Hacobian*

[*Manuscript received April 13, 1951*]

Summary

The A.C. polarography of uranium, iron, nickel, and manganese in suitable supporting electrolytes is discussed. Uranium was estimated in perchloric acid as the uranyl ion (UO_2^{++}) ; iron as the oxalato complex ion in neutral or acid oxalate. Nickel and manganese were estimated as cyano complexes which are reversibly reduced at the dropping mercury electrode.

I. Introduction

The A.C. polarography of heavy metals presents more difficulties than that of the metals discussed in previous papers (Breyer, Gutmann, and Hacobian 1950, 1951).

Thus, for example, Ni^{++} ion forms a firm hexaquo complex which is irreversibly reduced at the dropping mercury electrode ; the standard oxidation-reduction potential of the Fe^{+++}/Fe^{++} system is so positive ($+0 \cdot 51$ V. *v.* S.C.E.) that Fe^{+++} ions are reduced to Fe^{++} by the mercury ; the A.C. polarography of hexaquo Mn^{++} ion yields only small waves and is therefore unsatisfactory.

It is possible, however, to overcome most of the difficulties by proper choice of a suitable supporting electrolyte. Thus both nickel and manganese cyano complexes are reduced reversibly and consequently yield good A.C. waves in KCN as supporting electrolyte. In the case of iron, A.C. polarography can be carried out successfully by following the D.C. polarographic procedure recommended by Stackelberg and Freyhold (1940) and by Lingane (1946), i.e. by forming the stable oxalato complex.

The A.C. polarography of uranium presented no particular difficulties.

II. Uranium

On the evidence of Herasymenko (1925), Strubl (1938), and Kolthoff and Harris (1946, 1947), the D.C. polarography of the uranyl ion varies greatly with the pH and the nature of the supporting electrolyte. Plotting log $i_d - i/i$ against the applied potential E, a straight line with a slope of 60 mV. per logarithmic unit was obtained for uranyl ion in $0 \cdot 5$N perchloric and $0 \cdot 5$N sulphuric acids as supporting electrolytes. This confirms the original results observed by Herasymenko (1925) that the reduction of UO_2^{++} ion proceeds to the $+5$ valency state. Reduction of UO_2^{++} ions gave a well-defined A.C.

---

* Physico-Chemical Laboratory, Department of Agriculture, University of Sydney.

polarographic wave in the supporting electrolytes mentioned, and a practically linear relationship between ion concentration and maximum differential current was found. The calibration curve for UO_2^{++} in 0·5N perchloric acid and in the presence of air is shown in Figure 1.

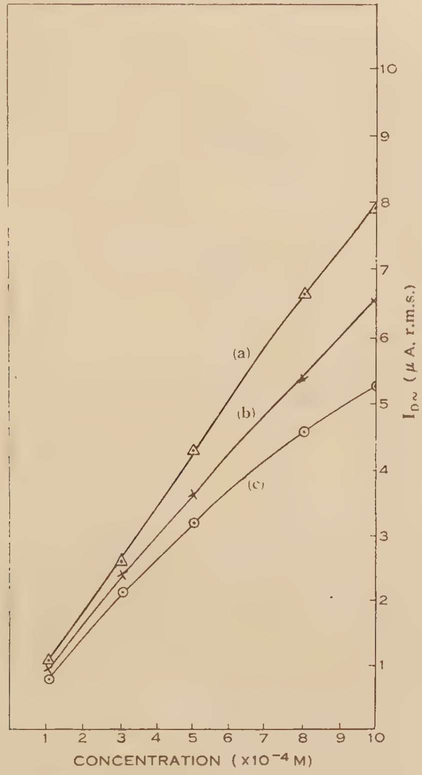


Fig. 1.—Calibration curves of Mn^{++}, UO_2^{++}, and Ni^{++} in the presence of air. A.C. 15 mV. r.m.s.; T, 20 °C.

(*a*) Mn^{++} in M KCN.
(*b*) UO_2^{++} in 0·5N H_2SO_4.
(*c*) Ni^{++} in M KCN.

III. Iron

Good A.C. waves were obtained with both ferric and ferrous ion in M oxalic acid and in M sodium or potassium oxalate as supporting electrolytes. Their summit potentials corresponded to the oxidation-reduction potential of the ferric-ferrous oxalato system. The characteristics of the D.C. polarograms due to the reduction of the oxalato ferric ion and to the oxidation of the corresponding ferrous complex ion have been studied by various authors (Stackelberg and Freyhold 1940; Toropova 1941; Lingane 1946) who have shown that the process is reversible.

The E_s value corresponded to the half-wave potential ($E_s = -0{\cdot}24$ V. *v*. S.C.E. in M potassium oxalate). In the presence of oxalic acid both the A.C.

and D.C. waves were shifted to more positive potentials as compared to those observed in neutral supporting electrolytes. The $E_{\frac{1}{2}}$ and E_s values in 0·5M oxalic acid (air-free) were +0·04 and +0·035 V. *v.* S.C.E. respectively. They varied with changes in pH. Ferrous and ferric ion gave identical A.C. polarograms in oxalic acid and in potassium oxalate as supporting electrolytes. Figure 2 shows the A.C. and D.C. polarograms of a mixture of Fe^{III} and Fe^{II} ions in potassium oxalate. The anodic part of the D.C. polarogram shows the oxidation of the ferrous complex whose concentration is represented by $(i_a)_d$ whereas the reduction of the ferric complex is represented by the cathode diffusion

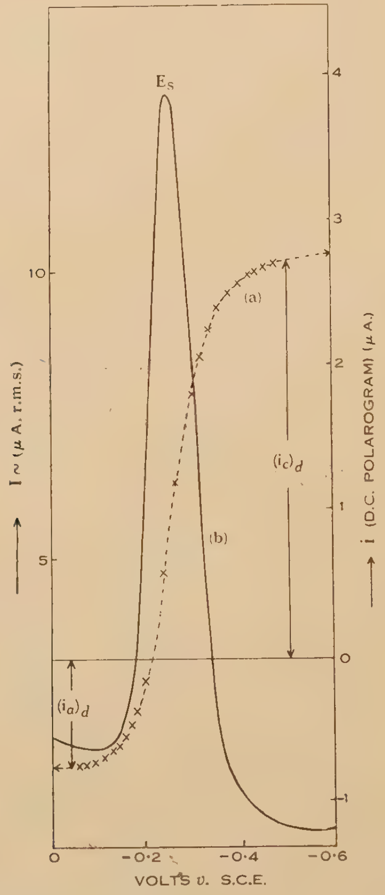


Fig. 2.—A.C. and D.C. polarograms of ferrous and ferric ion in neutral potassium oxalate as supporting electrolyte. A.C. 15 mV. r.m.s.; T, 20 °C.

(*a*) Combined cathodic-anodic D.C. polarogram of a solution of 0·5M neutral potassium oxalate (air-free) containing a mixture of ferrous and ferric ion (5×10^{-4}M).

(*b*) A.C. polarogram of (*a*).

current $(i_c)_d$ (Kolthoff and Lingane 1941, p. 168). The total iron concentration is thus represented by

$$i_d = (i_a)_d + (i_c)_d.$$

As expected, the summit potential of the A.C. wave practically coincides with the half-step potential of the combined anodic and cathodic D.C. polarogram. The calibration curve is given in Figure 3 which shows practically linear relationship between maximum differential current and the oxalato complex ion concentration.

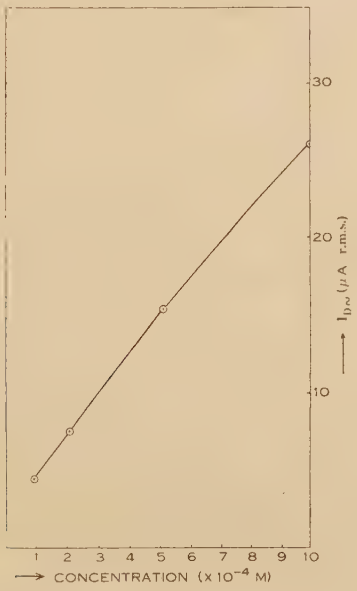


Fig. 3.—Calibration curve of iron in 0·5M neutral potassium oxalate and in the presence of air. A.C. 15 mV. r.m.s. ; T, 20 °C.

IV. NICKEL

A well-defined A.C. polarographic wave of nickel complex ion was obtained in a cyanide medium as supporting electrolyte. The observed summit potentials were about $-1 \cdot 40$ V. *v.* S.C.E. corresponding to the reduction of the tetracyano complex and in agreement with the $E_{\frac{1}{2}}$ values found by previous workers (Emelianova 1925 ; Kolthoff and Lingane 1941, p. 283). The value of the D.C. log plots (33 mV. per logarithmic unit for nickel cyanide complex in M KCN) indicates a two electron reduction process, in agreement with the results obtained by Kolthoff and Lingane (1941, p. 283).

Figure 1 shows the calibration curve of nickel ions in potassium cyanide solution, in the presence of air. The $i_{D\sim}$ values were practically linearly proportional to the ion concentration over a dilution range from 10^{-3} to 10^{-4} M.

The A.C. polarography of nickel was also carried out in buffered and unbuffered mixtures of pyridine and $0 \cdot 1$ M KCl solutions as supporting electrolytes. A.C. polarographic waves were observed at about −0·8 V. *v.* S.C.E. These waves were due to the reduction of the pyridinium complex ion (Lingane and Kerlinger 1941). They were not well defined. The $i_{D\sim}$ values were considerably less than those obtained in cyanide media.

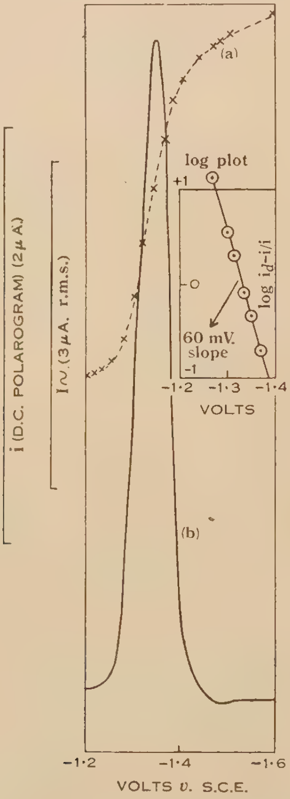


Fig. 4.—A.C. and D.C. polarograms of manganese ion in M potassium cyanide as supporting electrolyte (air-free).

(*a*) D.C. polarogram of 5×10^{-4}M manganese ion.

(*b*) A.C. polarogram of 5×10^{-4}M manganese ion.

V. MANGANESE

The A.C. and D.C. polarographic waves of manganese ions in KCN as supporting electrolyte are shown in Figure 4. The E_s value corresponded to the ordinary half-step potential (−1·34 V. *v.* S.C.E.).

According to Verdier (1939) the manganous cyanide complex in M KCN is reduced to the metal at the dropping mercury electrode. The present authors were unable to confirm Verdier's results. The D.C. step height obtained in the reduction of the manganous cyanide complex was less than half of that obtained in the reduction of the nickel complex. The reduction of the latter is a two electron process as has been shown by Kolthoff and Lingane (1941, p. 283) by comparing the diffusion current due to the $Ni(CN)_4^-$ ion in M KCN with that of thallous ion at the same molar concentration. The authors have repeated Kolthoff and Lingane's experiments with identical results. It follows, therefore, that the reduction of the manganous cyanide complex, under the conditions specified, does not lead to the metal but to the Mn^I state. This view is further corroborated by the result of the D.C. log plot as shown in Figure 4. In M KCN solution a practically linear relationship between manganous ion concentration and $i_{D\sim}$ values was observed in the concentration range from 10^{-4} to 10^{-3} M (cf. Fig. 1).

A.C. polarographic waves of manganese were also obtained in potassium chloride solution. The summit potential was identical with the D.C. half-step potential (−1·5 V. *v.* S.C.E.) as reported in the literature (Verdier 1939). The A.C. wave, however, is relatively small and not well suited for quantitative purposes.

VI. ACKNOWLEDGMENT

Our thanks are due to the Trustees of the Science and Industry Endowment Fund for a grant which enabled us to carry out this work.

VII. REFERENCES

BREYER, B., GUTMANN, F., and HACOBIAN, S. (1950).—*Aust. J. Sci. Res.* A **3** : 567.

BREYER, B., GUTMANN, F., and HACOBIAN, S. (1951).—*Aust. J. Sci. Res.* A **4** : 595.

EMELIANOVA, N. V. (1925).—*Rec. Trav. Chim.* **44** : 528.

HERASYMENKO, P. (1925).—*Chem. Listy* **19** : 172.

KOLTHOFF, I. M., and HARRIS, W. E. (1946).—*J. Amer. Chem. Soc.* **68** : 1175.

KOLTHOFF, I. M., and HARRIS, W. E. (1947).—*J. Amer. Chem. Soc.* **69** : 446.

KOLTHOFF, I. M., and LINGANE, J. J. (1941).—"Polarography." 1st Ed. (Interscience Publishers Inc.: New York.)

LINGANE, J. J. (1946).—*J. Amer. Chem. Soc.* **68** : 2448.

LINGANE, J. J., and KERLINGER, H. (1941).—*Industr. Engng. Chem.* (Anal. Ed.) **13** : 77.

STACKELBERG, M. v., and FREYHOLD, H. v. (1940).—*Z. Electrochem.* **46** : 120.

STRUBL, R. (1938).—*Coll. Trav. Chim. Tchécosl.* **10** : 466.

TOROPOVA, V. F. (1941).—*J. Gen. Chem. U.S.S.R.* **11** : 1211. (*Chem. Abstr.* **39** : 4016 (1945)).

VERDIER, E. T. (1939).—*Coll. Trav. Chim. Tchécosl.* **11** : 216.

POLAROGRAPHY WITH ALTERNATING CURRENTS

V. A.C. POLAROGRAPHY OF SOME COMMON ANIONS

By B. BREYER* and S. HACOBIAN*

[*Manuscript received April* 13, 1951]

Summary

The A.C. polarography of Cl^-, Br^-, CNS^-, OH^-, $SO_3^=$, CN^-, and $S^=$ is discussed. Chloride, bromide, hydroxide, and sulphide ions were estimated quantitatively by A.C. polarography in suitable supporting electrolytes.

The anodic A.C. waves present the following characteristic features :

(i) They are asymmetric and much sharper than the usual cathodic A.C. waves.

(ii) The summit potentials do not correspond to the relative half-step potentials but occur at voltages corresponding to the commencement of the anodic discharge.

(iii) In contrast to cathodic A.C. polarography the anodic A.C. electrode processes strongly interfere with one another ; the wave height of a given anion species is greatly reduced in the presence of anions dischargeable at more negative potentials.

I. INTRODUCTION

A comprehensive account of anodic D.C. polarography has been given by various workers in this field (Revenda 1934 ; Kolthoff and Lingane 1941, p. 322) who have shown that the relation between polarographic half-step potential ($E_{\frac{1}{2}}$) and anion activity is given by

$$E_{\frac{1}{2}}=K-\frac{0\cdot0591}{n}\log_{10}a_{\text{anion}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

where K is a constant, a activity, and n the valency of the anion. As can be seen, $E_{\frac{1}{2}}$ shifts to more negative values with increasing concentration (activity) of the anion.

The summit potentials of the A.C. polarographic waves also shift in accordance with equation (1) (cf. Table 1). But contrary to the results obtained in the A.C. polarography of cations, the summit potentials of the anodic A.C. waves were not identical with the corresponding half-step potentials. The difference between the respective E_s and $E_{\frac{1}{2}}$ values was quite appreciable in many cases as will be shown in the following sections.

II. THE HALOGEN IONS

The A.C. waves of the ions of the halogen series (with the exception of iodide ion) were well defined in a concentration range between 2×10^{-3} and 2×10^{-5} M. These waves were unsymmetrical about the equipotential line passing through the summit potential; the base current on the positive side of the summit potential was considerably depressed (Fig. 1).

---

* Physico-Chemical Laboratory, Department of Agriculture, University of Sydney.

The E_s values for Cl^- and Br^- ions did not coincide with the respective D.C. half-step potentials especially in the higher concentration ranges (cf. Table 1). A typical anodic A.C. wave of Cl^- ion (10^{-3} N) in perchloric acid as compared with its corresponding D.C. polarogram is shown in Figure 1. It is seen that the summit potential practically coincides with the beginning of the depolarization process.

TABLE 1

A.C. POLAROGRAPHY OF HALOGEN IONS

A.C. 15 mV. r.m.s.; T, 20 °C.

Ion	Supporting Electrolyte (air-free)	Ion Concn. ($N \times 10^{-3}$)	E_s *v.* S.C.E. (V.)	$E_{\frac{1}{2}}$ *v.* S.C.E. (V.)	*E v.* S.C.E. (calc. V.)	$i_{D\sim}$ (μA. r.m.s.)
Cl^-	0·5N $HClO_4$	0·1	+0·260	+0·290	+0·265	3·8
,,	,,	0·5	+0·223	+0·250	+0·225	25·5
,,	,,	1	+0·208	+0·24	+0·205	37
,,	,,	2	+0·1875	+0·235	+0·189	60
,,	0·5N $NaClO_4$	0·1	+0·262	+0·275	+0·265	3·5
,,	,,	0·5	+0·2285	+0·260	+0·225	21
,,	,,	1	+0·2135	+0·250	+0·205	35·5
,,	,,	2	+0·1945	+0·238	+0·189	49
Br^-	0·5N $NaClO_4$	0·1	+0·175	+0·23	+0·172	4·0
,,	,,	0·5	+0·1050	+0·14	+0·131	28
,,	,,	1	+0·0901	+0·125	+0·11	43
I^-	0·5N $NaClO_4$	1	—0·080* —0·220†	—‡	—0·095	—

* First I^- wave.

† Second I^- wave.

‡ Ill-defined D.C. polarogram.

The summit potentials and maximum differential current values were unaffected by the presence of air. The values for iodide ion are not given since its polarogram was not clearly defined. Iodide (10^{-3} N) in 0·5 N $NaClO_4$, however, gave a fairly good A.C. polarogram consisting of two waves; the summit potentials were —0·080 and —0·220 V. *v.* S.C.E. respectively.

Figure 2 shows the relation between $i_{D\sim}$ and anion concentration in perchloric acid, in the presence of air.

A.C. polarography seems particularly valuable for the quantitative estimation of Cl^- and Br^- since the A.C. waves of these ions are distinctly better defined than the corresponding D.C. steps. The waves were so pronounced and sharp that the summit potentials of 10^{-3} N solutions of these ions could be determined with an accuracy of better than ±0·5 mV. The A.C. waves of chloride and bromide ions strongly interfered with one another in the sense that the chloride wave height was considerably reduced in the presence of Br^- ions. At the same time the A.C. wave due to Br^- discharge increased somewhat in height.

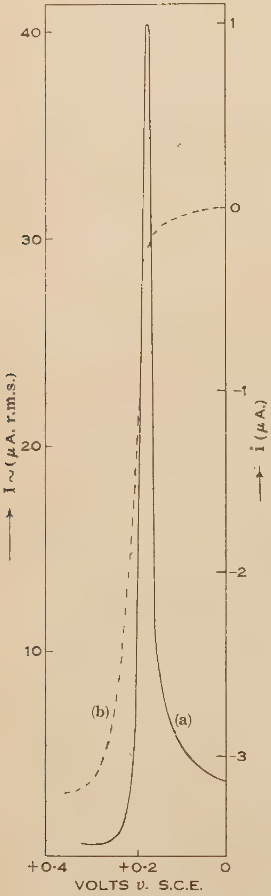


Fig. 1

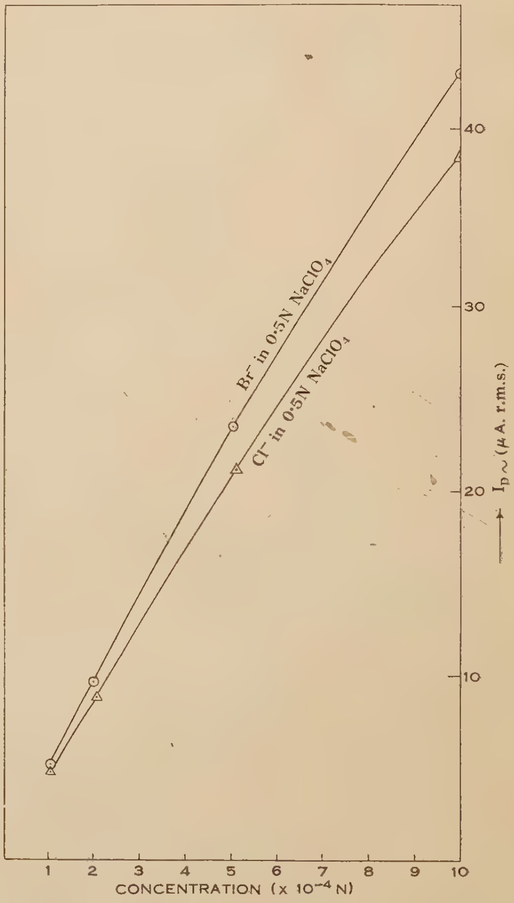


Fig. 2

Fig. 1.—Typical anodic A.C. polarogram of 10^{-3}N Cl^- in $HClO_4$ compared with the corresponding D.C. polarogram.

(*a*) A.C. polarogram of 10^{-3}N Cl^- in 0·5N $HClO_4$ as supporting electrolyte (air-free).

(*b*) D.C. polarogram of the same solution.

Fig. 2.—Calibration curves of Cl^- and Br^- in 0·5N $NaClO_4$ and in the presence of air. A.C. 15 mV. r.m.s.; T, 20 °C.

III. Thiocyanate, Hydroxide, and Sulphite Ions

The A.C. polarograms of the above ion species were not so well defined as those of Cl^- and Br^- but reasonably good waves were observed when the ion concentration exceeded 10^{-4} M. For $SO_3^{=}$, the polarographic estimation has to be carried out in the absence of air in order to prevent oxidation. Sodium perchlorate (0·5 M) was used as supporting electrolyte.

The summit potentials for 10^{-3} N CNS^-, OH^-, and $SO_3^=$ were +0·17, +0·14, and 0·00 V. *v.* S.C.E. respectively. As with halogen ions, the E_s values were more negative than the corresponding $E_{\frac{1}{2}}$ values (cf. Table 2).

TABLE 2

A.C. POLAROGRAPHY OF SOME COMMON ANIONS

A.C. 15 mV. r.m.s.; T, 20 °C.

Ion	Supporting Electrolyte (air-free)	Ion Concn. ($N \times 10^{-3}$)	E_s *v.* S.C.E. (V.)	$E_{\frac{1}{2}}$ *v.* S.C.E. (V.)	$i_{D\sim}$ (μA. r.m.s.)
OH^-	0·5N $NaClO_4$	0·01	+0·29	—	—
,,	,,	0·1	+0·285	—	1·0
,,	,,	0·5	+0·150	+0·14	3·5
,,	,,	1	+0·140	+0·15	7·0
$SO_3^=$	,,	0·5	0·00	+0·02	4·0
,,	,,	1	0·00	0·00	11·5
,,	,,	2	—0·005	—	18·5
CNS^-	,,	0·1	+0·22	—	3·0
,,	,,	1	+0·17	+0·20	21·5
,,	,,	2	+0·17	+0·195	38
$S^=$	N NaOH	0·1	—0·71	—0·71	1·3
,,	,,	0·5	—0·78	—0·785	11·2
,,	,,	1	—0·79	—0·795	20·9
CN^-	,,	0·1	—0·28	—0·28	1·0
,,	,,	0·5	—0·31	—0·31	6·9
,,	,,	1	—0·32	—0·325	13·4

The calibration curve for the hydroxide ion is shown in Figure 3. An approximate linear relationship between its concentration and maximum differential current was observed.

Thiocyanate and sulphite ions did not yield A.C. waves suitable for quantitative estimation. The waves were difficult to reproduce and the relationship between concentration and wave height was not linear.

IV. CYANIDE AND SULPHIDE IONS

Good A.C. waves of CN^- and $S^=$ were obtained in strong alkaline supporting electrolytes (e.g. N NaOH). The summit potentials for 10^{-3} N CN^- and $S^=$ were —0·32 and —0·79 V. respectively (*v.* S.C.E.). Figure 3 shows the relationship between $i_{D\sim}$ values and ion concentration of sulphide and cyanide ions in N NaOH, in the presence of air.

V. CORRELATION BETWEEN SUMMIT POTENTIALS OF Cl^- AND Br^- AND THE SOLUBILITY PRODUCTS OF MERCUROUS CHLORIDE AND BROMIDE

The potential of the dropping electrode at any point of the D.C. polarographic wave is expressed by equation (**2**)

$$E = E^0_{Hg} + \frac{0 \cdot 0591}{2} \log_{10} [Hg_2^{++}], \quad \ldots\ldots\ldots\ldots \quad (2)$$

where E^0_{Hg} is the standard potential of the reaction shown in the equation

$$2Hg \leftrightharpoons Hg_2^{++} + 2e. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

In the presence of halogen ions the critical concentration of mercurous ions at the electrode surface, $(C^0_{Hg_2}{}^{++})$, is determined by the solubility of mercurous halide produced and the concentration of halogen ions in the bulk of the solution.

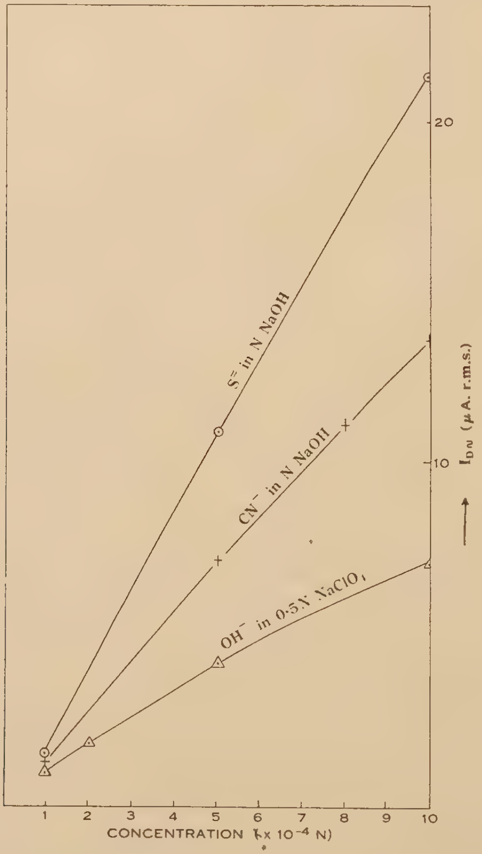


Fig. 3.—Calibration curves of $S^=$, CN^-, and OH^- in the presence of air. A.C. 15 mV. r.m.s. ; T, 20 °C.

The standard potential of the system $Hg_2^{++}/2Hg$ is $+0 \cdot 553$ V. *v.* S.C.E. Substituting this and the value of $C^0_{Hg_2}{}^{++}$ in equation (**2**), the values of E corresponding to the commencement of the anodic discharge can be calculated, as pointed out by Kolthoff and Lingane (1941, p. 322). The anodic D.C. step begins when the electrode is made sufficiently positive so that the concentration of mercurous ions produced at the electrode surface exceeds the solubility product of mercurous halide. A close agreement between the calculated E values and the observed summit potentials (E_s) of the differential wave was obtained for chloride ions as shown in Table 1. For bromide ion the agreement was good up to a concentration of 5×10^{-4} N.

VI. ACKNOWLEDGMENT

The authors are much indebted to the Science and Industry Endowment Fund for financial assistance.

VII. REFERENCES

KOLTHOFF, I. M., and LINGANE, J. J. (1941).—"Polarography." 1st Ed. (Interscience Publishers Inc.: New York.)

REVENDA, J. (1934).—*Coll. Trav. Chim. Tchécosl.* **6**: 453.

DIELECTRIC ABSORPTION IN SOLID ALIPHATIC METHYL ESTERS

By J. S. Dryden* and H. K. Welsh†

[*Manuscript received April 24, 1951*]

Summary

The dielectric absorption in three aliphatic methyl esters, methyl palmitate, methyl stearate, and methyl behenate, has been measured over a wide frequency range at temperatures of 25, 2, and —18 °C. The measurements reveal that there are two absorptions in these esters, both arising from relaxation mechanisms.

Deductions made from the results indicate that one of the absorptions is due to the rotation of the ester molecule from one position of equilibrium in the crystal lattice to another, and that the other absorption arises from rotation of the polar group (which in these molecules is at one end) from one position of relative equilibrium to another.

I. Introduction

It is known that the aliphatic long-chain ester, methyl palmitate, possesses at least two regions of dielectric absorption (Dryden and Jackson 1948 ; Jackson 1949) with the maxima occurring in the radio frequency spectrum. It was suggested (Dryden and Jackson 1948) that the lower frequency absorption is associated with a rotation of the whole molecule, and that the higher frequency absorption may be due to a resonance effect. An alternative explanation is that the latter absorption is associated with a rotation of the end polar group, or a part thereof, from one position of equilibrium to another.

The work reported in the present paper was undertaken to elucidate the mechanisms of energy absorption involved in this type of compound. It was considered that measurements on a series of methyl esters of differing chain-lengths and at different temperatures would provide crucial information. Consequently, the dielectric properties of the methyl esters, methyl palmitate ($C_{15}H_{31}COOCH_3$), methyl stearate ($C_{17}H_{35}COOCH_3$), and methyl behenate ($C_{21}H_{43}COOCH_3$) were investigated over a wide frequency range at ambient and lower temperatures.

II. Materials

Methyl stearate and methyl palmitate were prepared by esterification of commercial stearic acid, containing approximately 40 per cent. palmitic acid, with excess methanol in the presence of concentrated sulphuric acid. The esters were separated by fractional vacuum distillation and recrystallized from methanol several times until tan δ at 50 c/s. was not greater than $0 \cdot 002$. Methyl behenate was prepared by the action of diazomethane on a solution of behenic acid in dry

---

* Division of Electrotechnology, C.S.I.R.O., University Grounds, Sydney.

† Division of Industrial Chemistry, seconded to the Division of Electrotechnology, C.S.I.R.O., University Grounds, Sydney.

ether. The ether was distilled from the ester, which was recrystallized from methanol as for the other esters. These compounds had the following melting points :

Methyl palmitate	30·5–30·7 °C.
Methyl stearate	38·6–39·2 °C.
Methyl behenate	50·6–51·0 °C.

III. METHODS OF MEASUREMENT

Five methods of measurement were used in the frequency range covered. Measurements at frequencies from 200 c/s. to 100 kc/s. were carried out by means of a bridge with four capacitive arms, using a guard ring electrode assembly. Between frequencies of 90 kc/s. and 50 Mc/s., a series resonant circuit was used. Measurements at frequencies of 100, 150, and 300 Mc/s. were obtained with a resonant system developed by Hollway and Cassidy (unpublished), similar to that described by Hollway (1940). In this system a short length of coaxial line is used as the inductive element. A standing-wave method of measuring dielectric properties (Roberts and von Hippel 1946) was used at frequencies of 1000, 3000, and 8600 Mc/s. An H_{01} resonator, similar in construction to that described by Bleaney, Loubser, and Penrose (1947), was used for measurements at 24,000 Mc/s.

Measurements at temperatures of 2 and −18 °C. at frequencies of 1000, 3000, and 8600 Mc/s. were obtained by working in rooms maintained at these temperatures,* as were measurements at 2 °C. and for frequencies between 90 kc/s. and 50 Mc/s. Otherwise, measurements at the lower temperatures were obtained by means of a cold, dry air stream which circulated in or around the respective resonator or electrode assembly. Care was taken to prevent ice formation on parts of the electrode assemblies which were in the electric field.

The accuracy of measurement is considered, in general, to be ±2 per cent. for ε' and ±5 per cent., or 0·0003, whichever is the greater, for $\tan \delta$ at room temperature ; but to be slightly worse at the lower temperatures. At −18 °C. the measurements of $\tan \delta$ are considered to be accurate to ±10 per cent.

IV. PREPARATION OF SPECIMENS

The specimens required for the different measuring techniques were prepared by pressing powdered material in chromium-plated brass moulds of suitable size and shape. The densities of the samples were measured and found to be 0·97±0·01 g./cc. From the available crystallographic data, the crystal density is estimated to be 1·00 g./cc.

V. RESULTS

The measured $\tan \delta$ values at 25 °C. for the three esters are shown plotted as a function of frequency in Figure 1. Figures 2, 3, and 4 show $\tan \delta$ as a function of frequency at each temperature for methyl palmitate, methyl stearate, and methyl behenate respectively. Relative permittivity values are not shown; the values obtained decrease continually, within experimental error, with

---

* The authors wish to thank Dr. J. R. Vickery, Chief of the Division of Food Preservation and Transport, C.S.I.R.O., for this facility.

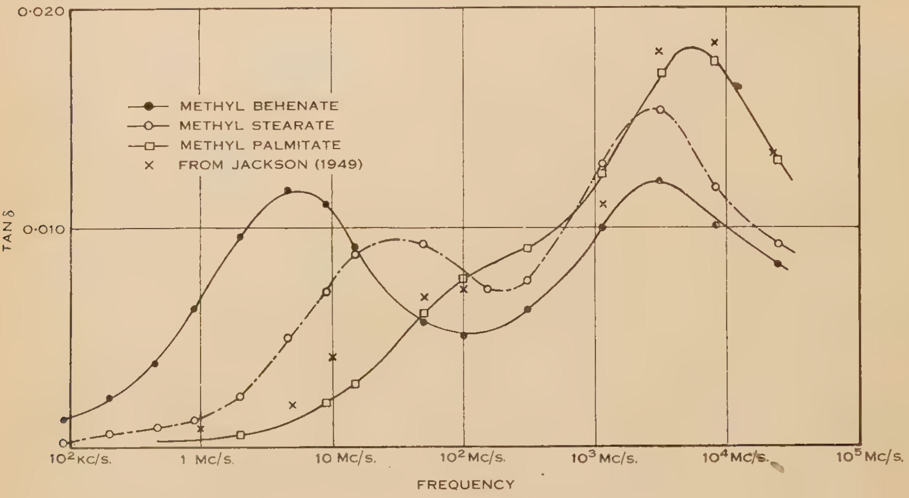


Fig. 1.—Dielectric loss of methyl esters at 25 °C.

increasing frequency. For methyl behenate, for example, the relative permittivity (ε′) at room temperature decreases from 2·60 at 100 kc/s. to 2·46 at 24,000 Mc/s.

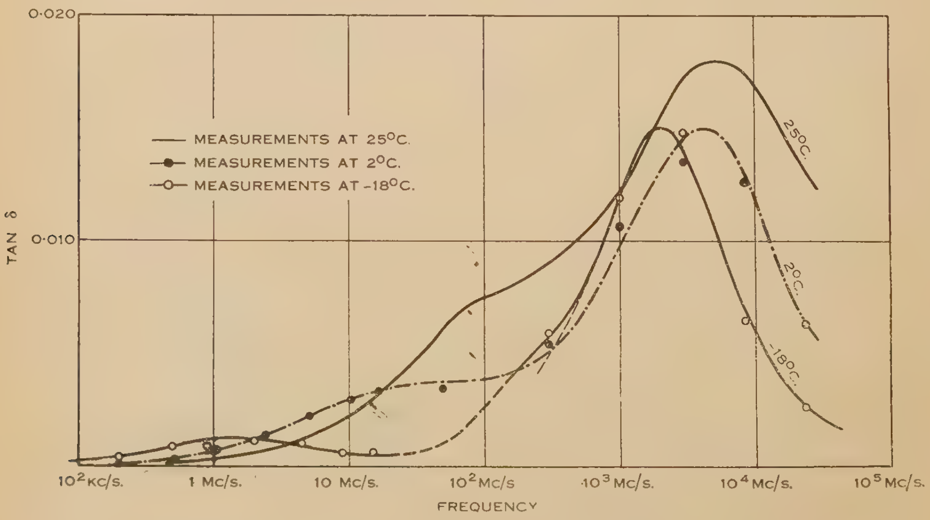


Fig. 2.—Dielectric loss of methyl palmitate.

Because the errors in the measurement of relative permittivity are of the same magnitude as the total variation in this quantity, relative permittivity values will not be used in the discussion. The observed values, however, are consistent with the explanations suggested below.

It can be seen that there are two distinct regions of energy absorption in these methyl esters. The lower frequency absorption is markedly dependent on chain-length; at any one temperature the frequency at which maximum absorption occurs decreases with increasing chain-length of the ester. On the other hand, in the higher frequency absorption, the frequency at which maximum tan δ occurs does not depend to anything like the same extent on the length of the ester molecule. In fact, while it is apparent that this absorption occurs at a slightly higher frequency in methyl palmitate, the measurements suggest that the frequencies of maximum absorption in the other two esters only differ by an amount which is within the experimental error.

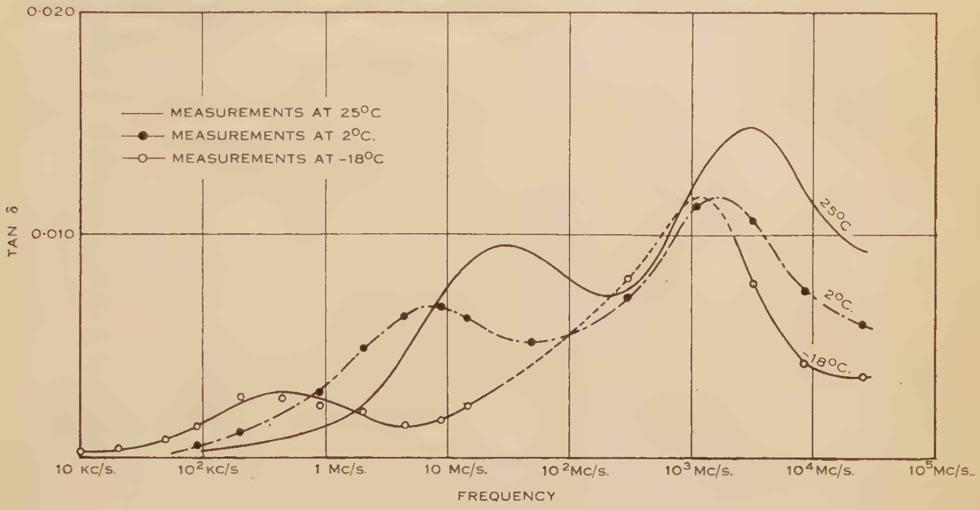


Fig. 3.—Dielectric loss of methyl stearate.

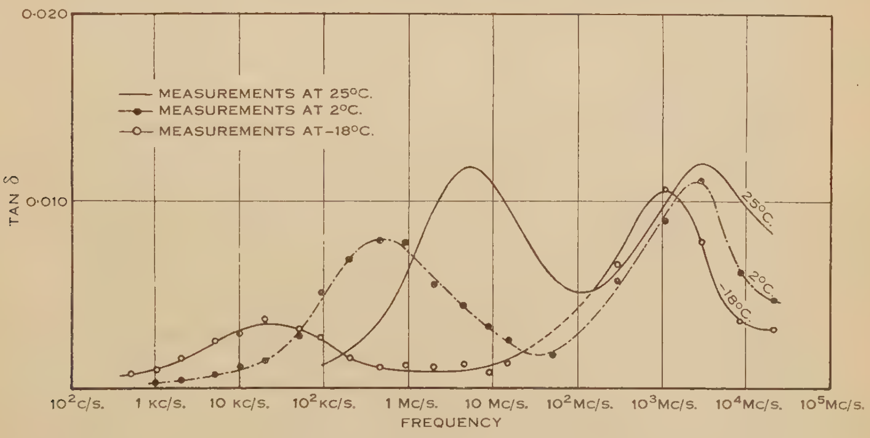


Fig. 4.—Dielectric loss of methyl behenate.

Included in Figure 1 are the values of tan δ reported by Jackson (1949) for methyl palmitate, at a temperature of 22 °C. Agreement between the two sets of measurements is satisfactory when it is remembered that there is a temperature difference of 3 °C. and that the lower frequency absorption is appreciably dependent on temperature. Relative permittivity values also agree to within 2 per cent. The agreement between the measurements in these two samples of methyl palmitate, coupled with the fact that during the course of this investigation the esters were recrystallized several times without any alteration in dielectric properties being observed, would suggest that it is unlikely that any of the measured absorption is due to impurity.

In the initial measurements (Dryden and Jackson 1948) a dielectric loss was obtained at audio frequencies, increasing in magnitude with decreasing frequency, and this dielectric loss was assumed to be a property of the pure material. Subsequent measurements reported by Jackson (1949), on some of the same material further purified by recrystallization, showed that this loss decreased with each recrystallization. It is now known that this dielectric loss at low frequencies is possessed by many solid aliphatic compounds which have been separated and purified by vacuum distillation (Hamon and Meakins 1950). However, with most compounds the magnitude of the loss decreases with each recrystallization, from which it is concluded that this low-frequency dielectric loss must be associated with impurities.

VI. Discussion

(a) *Crystal Structure*

The detailed crystal structure of methyl esters, or indeed of any simple polar long-chain aliphatic compound, has not been determined. However, the basic crystal structure of this type of compound is that of the paraffinic hydrocarbons, an analysis from X-ray diffraction measurements of one of which is available (Müller 1928). The carbon atoms in each molecule form a plane zigzag, the unit cell being a rectangular prism containing four molecules. These four molecules lie in the unit cell in the manner shown in Figure 5, the base dimensions of the unit cell being $a=7\cdot5$ Å and $b=5\cdot0$ Å. The carbon-carbon chains are perpendicular to this base plane, the third axis c being slightly greater than twice the chain-length measured between the centres of the end carbon atoms. In Figure 5, the dotted lines represent molecules in a layer below that to which the solid lines refer, these molecules being slightly displaced in such a way that their ends pack into the spaces between the molecules of the upper layer. X-ray diffraction measurements reveal, however, that the carbon-carbon chains are not arranged perpendicularly to the base plane in all paraffinic compounds. With certain assumptions (e.g. MacArthur 1944), it is possible from the increment in long spacing and the calculated addition to the length of the molecule, to estimate the angle of tilt in an homologous series.

It has been concluded by Malkin (1931) that the stable crystalline form of methyl esters contains double molecules, i.e. the polar groups of molecules in successive layers are adjacent to one another. He estimated that, in methyl esters containing an odd number of carbon atoms, as do the esters used in these experiments, the angle of tilt is 63°. Malkin found that the lateral spacings were not greatly different to those of the hydrocarbons. It is most probable that in the lowest energy state each molecule is extended as in Figure 6, in which the arrangement of the atoms in the polar group causes the least distortion from the spacings found in the hydrocarbons. The experimental results will be discussed on the basis of this crystal structure.

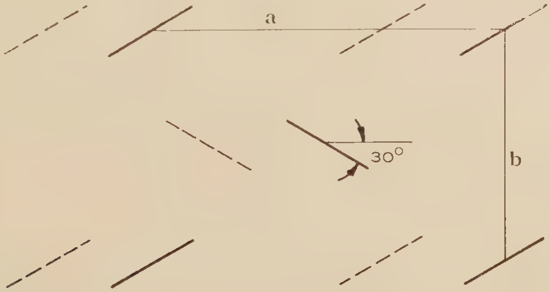
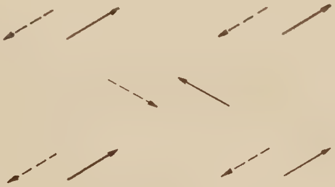


Fig. 5

(*b*) *Theoretical*

From the crystal structure it can be seen that each molecule has two positions of relative equilibrium, the molecule moving from one position into the other by a corkscrew rotation through 180° (Fig. 8 (*a*)). The dielectric properties



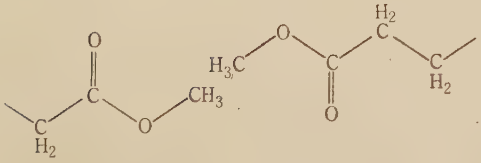


Fig. 6

of a model which is similar to that postulated here, has been discussed by Fröhlich (1949). The dielectric absorption associated with such a transition can be characterized by a relaxation time (τ) the value of which is determined by the height (H) of the potential energy barrier

$$\tau = \frac{2\pi}{\omega_a} A e^{H/kT}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

(k is Boltzmann's constant, T the absolute temperature) where π/ω_a is the average time required by an excited molecule to turn from one position to the other and A is a factor which must be introduced to account for the fact that a molecule is capable of internal excitation. Provided the value of the energy barrier (H) is identical for each molecule, then tan δ as a function of frequency will be described by a Debye curve $\tan\delta = B[\omega\tau/(1+\omega^2\tau^2)]$, where ω is the angular frequency and B is independent of frequency. The frequency of maximum absorption (f_{max}) is equal to $1/2\pi\tau$.

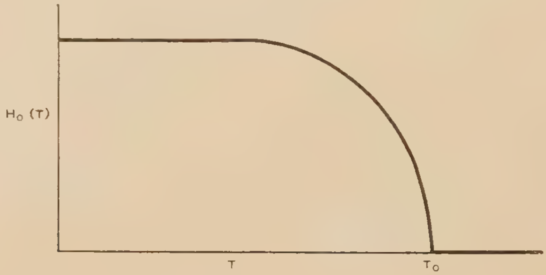


Fig. 7

The decrease in relative permittivity ($\varepsilon_s - \varepsilon_\infty$) in passing through an absorption region and also the magnitude of the energy absorption is dependent on H_0, the average energy difference between the two positions of equilibrium. ($\varepsilon_s - \varepsilon_\infty$) is proportional to

$$\frac{1}{kT} \times \frac{1}{(1+\cosh H_0/kT)}, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

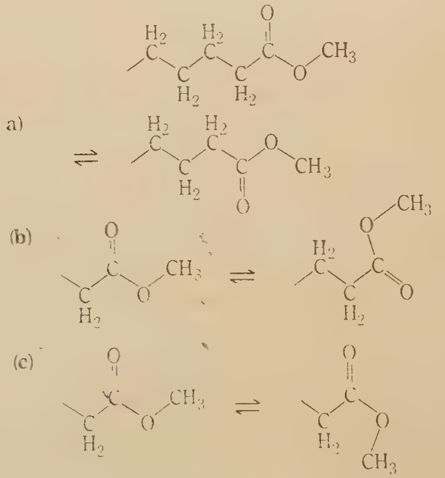


Fig. 8

which approximates to $(1/kT)e^{-H_0/kT}$ for values of $H_0 >> kT$. Fröhlich (1949) suggests that the energy difference H_0 varies with temperature in the manner shown in Figure 7. For temperatures greater than T_0 and for temperatures less than that at which H_0 becomes essentially the same as that of absolute zero, Fröhlich shows that the absorption should be characterized by a single relaxation time, i.e. a Debye curve should describe the variation of tan δ with frequency. However, at those temperatures where $H_0(T)$ is varying with temperature the absorption will not be characterized by a single relaxation time.

In addition to a rotation involving the complete molecule, the end polar group may be able to move into other positions of equilibrium as suggested in Figures 8 (*b*) and 8 (*c*). This assumes that the planar zigzag structure of the –C–C– chain persists and that there are no positions of equilibrium in which the –C–C– skeleton is twisted.

(*c*) *Lower Frequency Absorption*

The fact that this lower frequency absorption is markedly dependent on the chain-length of the ester, suggests that this absorption should be identified with rotation of the complete molecule from one position of equilibrium to another. Further confirmation of this is obtained from an examination of the variation of the frequency of maximum absorption with temperature. The frequencies of maximum absorption for methyl behenate and for methyl stearate are given in Table 1.

Owing to the decreased magnitude of the absorption, there are greater errors in the estimation of the frequency at which maximum absorption occurs at −18 °C., than at the higher temperatures. No values are given for methyl palmitate since it is difficult to extract this information because of the overlap between the two absorptions.

Table 1

Temperature (°C.)	Frequency of Maximum Tan δ	
	Methyl Behenate	Methyl Stearate
25	5·5 Mc/s.	35 Mc/s.
2	500 kc/s.	7 Mc/s.
—18	20 kc/s.	500 kc/s.

It can be seen from Table 1 that the variation of log (f_{max}) with $1/T$ is greater for methyl behenate than for methyl stearate. This means (see equation (1)) that the mean energy barrier to rotation is greater for the compound with the longer hydrocarbon chain, a result to be expected if this absorption arises from a rotation of the complete molecule. Using equation (1), approximate values for the mean energy barrier to rotation can be obtained; these are 17 kcal./mol. (28 kT for $T=300$ °K.) for methyl behenate and 11 kcal./mol. for methyl stearate.

The other quantity, in addition to the height of the potential barrier, which is significant in a relaxation absorption, is the average energy difference (H_0) between the two equilibrium positions. As discussed above, the magnitude of the dielectric absorption, or more properly, the area under the tan δ v. log (frequency) curve, is proportional to $e^{-H_0/kT}$ for sufficiently high values of H_0 (equation (2)). An approximate calculation gives, for methyl behenate, 2·9 kcal./mol. as the mean value of H_0 between 0 and 25 °C.; this value of H_0 is sufficiently high to justify the use of the above approximation. H_0 is, however, not constant within the temperature range covered by this investigation (−20 to 25 °C.). This can be seen from an inspection of Figures 2, 3, and 4, which shows that the expression log ($T \times \tan \delta_{max}$) does not vary linearly with $1/T$ as required by equation (2) if H_0 were constant. In the case of methyl stearate the maximum values of tan δ are in the ratio of 1·0 : 0·8 : 0·3 for temperatures of 25, 2, and −18 °C. respectively. The energy difference between the two positions of equilibrium is therefore increasing with decreasing temperature and suggests that, for these esters, the temperature range used in this investigation corresponds to a part of the curved region of H_0 as a function of temperature shown in Figure 7. Fröhlich (1949) suggests that for the temperatures in this region the absorption will be broader than a Debye curve, as is the case for these esters.

(d) *Higher Frequency Absorption*

As mentioned in the introduction, the suggested mechanisms for this higher frequency absorption were (i) a resonance absorption, or (ii) a relaxation mechanism. Fröhlich (1945) has predicted, on theoretical grounds, that a resonance absorption may be detected in this type of compound at frequencies between 3×10^5 and 3×10^6 Mc/s.

A means of differentiating between a resonance and a relaxation absorption is by an examination of the manner in which the absorption changes with temperature. In the case of a resonance absorption, the frequency at which maximum absorption occurs should be almost independent of temperature, the only change being a sharpening of the absorption with decreasing temperature. For a relaxation absorption, on the other hand, the absorption curve should move to lower frequencies with decreasing temperature. Consequently, an immediate conclusion from the variation of the dielectric loss with temperature (Figs. 2, 3, and 4) for all three compounds is that this absorption must arise from a relaxation and not from a resonance mechanism.

Using equation (1), it is possible to calculate an approximate value of the activation energy from the frequencies of maximum absorption. The value of H obtained (5 kcal./mol.) is the same for methyl palmitate and for methyl behenate, within the error of this determination. This is confirmation of the suggestion that the absorption is due to the rotation of an end polar group, since this energy barrier (H) should be approximately the same irrespective of the length of the hydrocarbon chain attached to the end group. (The numerical values obtained for $A\pi/2\omega_a$ are 3×10^{-15} second for methyl palmitate and 5×10^{-15} second for methyl behenate.)

For methyl behenate and for methyl stearate, the measurements at a frequency of 1000 Mc/s. at −18 °C. suggest that there is another absorption with a maximum at still higher frequencies. Allowing for this, the absorption under discussion in this section is not much wider than that represented by a Debye curve. Also, it was found that a Debye curve fits the measurements of methyl palmitate at −18 °C. except for some divergence on the lower frequency side, and the curve shown in Figure 2 is such a curve. A comparison can be made between the magnitudes of the absorption in different esters. From the −18 °C. measurements, the ratio of the maxima in methyl palmitate and methyl behenate is 1·4 : 1 which is in good agreement with the ratio 1·33 : 1 of the dipole concentration per unit volume.

Because of the absence of a measurement on methyl stearate at −18 °C., it is difficult to extract information from the measurements on this ester. However, the measurements are consistent with the suggestion that the absorption in methyl stearate has a maximum at each temperature, at approximately the same frequency as for methyl behenate but about 15 per cent. greater in magnitude. From the ratios of dipole concentration in these esters, the dielectric absorption would be expected to be 20 per cent. higher in the stearate.

The deductions made from the experimental evidence are thus consistent with the suggestion that this energy absorption arises from a relaxation mechanism associated with the rotation of an end polar group. The two possible rotations from one position of equilibrium to another are shown in Figures 8 (*b*) and 8 (*c*). It is not possible to state which mechanism is involved in this absorption but the authors consider that the rotation of the CH_3 group about an axis along the C–O bond from one position of equilibrium to another is the more likely. The basis for this is the deduction of a value for H_0 (using equation (**2**)) of the order of kT (for $T=0$ °C.).

Cook and Buchanan (1950) have published some measurements on methyl palmitate at frequencies between 600 and 10,000 Mc/s. and at temperatures of 22, 0, −20, and −80 °C. A comparison of their measurements with those reported in this paper indicate satisfactory agreement at temperatures of 22 and 0 °C. At −20 °C. the value of tan δ given by Cook and Buchanan at 3000 Mc/s. is approximately 25 per cent. less than that shown in Figure 2. These authors consider that their results are consistent with the existence of a resonance absorption, but there is nothing in their results that conflicts with the conclusions reached above. The measurements of Cook and Buchanan at −80 °C., together with a measurement at the same temperature at a frequency of 50 Mc/s. made by the present authors, indicate that at −80 °C. the magnitude of the absorption in methyl palmitate has considerably decreased from its magnitude at −20 °C.

(*e*) *General*

The measurements discussed above were carried out on material prepared by crystallization from solution. Some measurements were made with specimens prepared from solid which had been cooled from the melt and powdered. It was found that this latter material had dielectric properties different from the material crystallized from solution. The magnitude of the lower frequency absorption was greater and the magnitude of the higher frequency absorption less for the melted esters. The dielectric loss was, however, dependent on time and eventually reached the same value, at each frequency, as for the solid crystallized from solution. Presumably in cooling from the melt, some of the ester crystallized in a metastable form and this metastable form changed with time into the stable form, which is in agreement with the observation of Phillips and Mumford (1934) that methyl esters exhibit monotropic dimorphism.

The above discussion of the two regions of dielectric absorption in methyl esters suggests that the higher frequency absorption should be peculiar to compounds where the ester group is near the end of the molecule. Consequently some measurements were carried out on cetyl palmitate ($C_{15}H_{31}COOC_{18}H_{37}$) in which the ester group is near the middle of the molecule. Cetyl palmitate has an absorption with a maximum at 600 kc/s., but apart from this, no further absorption was detected up to 10^4 Mc/s.

VII. Acknowledgments

The authors wish to thank Mr. F. J. Lehany for his help and encouragement during the course of this investigation, and Mr. M. C. McGregor for his assistance in the construction of apparatus.

VIII. References

Bleaney, B., Loubser, J. H. N., and Penrose, R. P. (1947).—*Proc. Phys. Soc.* **59** : 185.
Cook, H. F., and Buchanan, J. T. (1950).—*Nature* **165** : 358.
Dryden, J. S., and Jackson, W. (1948).—*Nature* **162** : 656.
Fröhlich, H. (1945).—B.E.A.I.R.A. Rep. L/T 157.
Fröhlich, H. (1949).—" Theory of Dielectrics." (Oxford Univ. Press.)
Hamon, B. V., and Meakins, R. J. (1950).—*Nature* **166** : 29.
Hollway, D. (1940).—*J. Instn. Engrs. Aust.* **12** : 79.
Jackson, W. (1949).—*Nature* **164** : 486.
MacArthur, I. (1944).—*Proc. Leeds Phil. Soc.* **4** : 170.
Malkin, T. (1931).—*J. Chem. Soc.* **1931** : 2796.
Müller, A. (1928).—*Proc. Roy. Soc.* A **120** : 437.
Phillips, J. W., and Mumford, S. (1934).—*J. Chem. Soc.* **1934** : 1657.
Roberts, S., and Hippel, von A. (1946).—*J. Appl. Phys.* **17** : 610.

THE ADSORPTION OF SODIUM CETYL SULPHATE ON CASSITERITE

By G. R. Edwards* and W. E. Ewers*

[*Manuscript received June 29, 1951*]

Summary

The adsorption of sodium cetyl sulphate on pure synthetic cassiterite has been measured under various conditions of concentration and pH and the results related to flotation. It was found that only a small proportion of the surface need be covered with collector ions to make flotation possible. At high concentrations of collector no evidence was found of a double layer of collector ions on the mineral surface, and it is concluded that the cessation of flotation at these concentrations is due to " armouring " of the bubble surfaces.

A mechanism for the adsorption of cetyl sulphate ions is proposed, in which these ions exchange with hydroxyl ions on the hydrated surface. An attempt is made to describe the ionic equilibria at the cassiterite surface.

Samples of natural cassiterite differed from one another and from the synthetic mineral in their ability to be floated under certain conditions : this is attributed to the effect of metallic impurities.

I. Introduction

It is now generally recognized by workers on the flotation of minerals, that, in order to make mineral particles hydrophobic so that they may become attached to bubbles of air, it is necessary to adsorb a suitable reagent on the mineral surface. The reagents or " collectors " used most commonly are alkali xanthates for sulphide minerals, and soaps or long-chain amines for minerals other than sulphides. All these reagents are amphipathic compounds possessing a polar group through which, it is believed, the link is made to the adsorbing surface, and a non-polar group which composes the new hydrophobic outer surface of the mineral particle. Several workers have attempted to measure the quantity of collector adsorbed as a function of collector concentration, their aims being variously, to determine what proportion of the surface must be covered to allow flotation to occur, to determine why flotation ceases at high concentrations of long paraffin chain reagents, or more generally to investigate the process of adsorption.

Taggart and Hassialis (1946) showed that an extremely low concentration (about 0·0007 mg./l.) of potassium ethyl xanthate was sufficient to produce a hydrophobic galena surface of contact angle 38°. Cook and Nixon (1950) calculated from these data that only 3·6 per cent. of the surface could have been covered. Using galena and potassium ethyl xanthate, Gaudin and Preller (1946) found that with about 50 per cent. of the surface of galena covered the recovery of galena in a flotation cell was 75 per cent. Simard, Chupak, and

---

* Division of Industrial Chemistry, C.S.I.R.O., Melbourne.

Salley (1950) reported divergent results for the amount of di-*sec.*-butyl dithiophosphate necessary to give 70 per cent. recovery of galena. In one test only 1 per cent. coverage was required and in another 14 per cent. Their determinations of surface area, however, were not reliable.

Gaudin and Bloecher (1950) attempted to measure the adsorption of dodecylamine on samples of quartz, the surface area of which had been determined by nitrogen and krypton adsorption. They found that 80 per cent. of the quartz could be floated when only 3 per cent. of the surface was covered with dodecylamine.

The cessation of flotation at high concentrations of collector was studied by Held and co-workers (1935, 1936). They advanced evidence that a double layer of collector ions is formed, as in a lamellar micelle, with the hydrophilic polar groups outermost. They determined the adsorption isotherms for sodium laurate from aqueous solution and for lauric acid from benzene solution on barium sulphate. The isotherm in each system was of the same form, the amount adsorbed approaching a constant maximum value as the concentration was increased. For the soap solution, however, this maximum corresponded to more than twice the maximum amount of lauric acid adsorbed which was assumed to represent a closely packed monolayer. Similar results were obtained using cinnabar and, as collectors, sodium oleate and sodium nonylate. Held's explanation of the phenomenon received support from Gaudin and Rizo-Patron (1943) and Taggart and Arbiter (1943). Gaudin and Bloecher (1950) have concluded that multilayer adsorption is possible in the system quartz-dodecylamine acetate, but their result must be regarded with suspicion since they did not account for adsorption of collector on the fritted glass filter used in the tests.

Wark (1936) suggested that the contact between bubble and mineral might be hindered also by an " armouring " of the bubble surface by adsorbed collector molecules. This view is supported by the continuously increasing induction period for contact on quartz and paraffin wax observed by Wark as the concentration of cetyl trimethyl ammonium bromide is increased. In the same paper this author showed, however, that the contact angle on a paraffin wax surface decreased continuously from 107° to zero as the concentration of cetyl trimethyl ammonium bromide was increased. A change at the air-water interface alone could not have produced this effect.

In a third explanation Dean and Ambrose (1944) proposed that at high concentrations of paraffin chain collectors, the hydrophobic film is deterged from the surface. This theory cannot be reconciled with Held's results and was severely criticized by Rogers and Sutherland (1946).

In the present investigation we sought to obtain unequivocal evidence, with respect to a particular system, concerning the two problems noted above and also to elucidate the nature of the adsorption process. The system chosen was cassiterite and sodium cetyl sulphate (Na$\overline{\text{CS}}$). Cassiterite is chemically inert, and is insoluble in the sense that the concentration of its constituent ions in the solution is negligible. It can be prepared in a very pure form and as a crystalline non-porous powder of high specific surface. In one or more of these respects it is superior to other minerals which have been chosen for similar studies. Sodium cetyl sulphate possesses the advantages that it is stable in aqueous solution, that its solutions are unaffected by carbon dioxide, and that a sensitive analytical method is available for its determination at low concentrations.

The flotation of cassiterite using sodium cetyl sulphate as collector has been studied by Hergt, Rogers, and Sutherland (1946).

II. MATERIALS

(i) *Cassiterite.*—Cassiterite from four sources was used during the investigation. For adsorption measurements it was necessary to obtain a sample of high specific surface. Stream cassiterite (cassiterite *A*), from an unknown Australian location, was treated with zinc and hydrochloric acid to reduce the surface to metallic tin. This procedure facilitated hand-picking to avoid visible inclusions. The hand-picked material was ground wet and elutriated in distilled water to give the fraction less than 20 μ. The extreme slimes were removed by decantation after centrifuging for 10 minutes at 900 × g. The sample so obtained was seriously contaminated with silica from the porcelain mill. Silica was removed with hydrofluoric acid, yielding a sample which contained 99·3 per cent. SnO_2. It was slightly off-white in colour. Its surface area, determined by nitrogen adsorption, was 2·45 m.^{2}/g.

Some adsorption tests were made with the porcelain-milled sample, but it was decided to continue the measurements using synthetic cassiterite so that the possibility of activation by foreign metal ions in the mineral could be avoided. The preparation of this material (cassiterite *B*) from "Analar" grade tin is described in Appendix I. The surface area of this sample, from nitrogen adsorption, was 10 m.^{2}/g. By chemical analysis it contained 100·0 ± 0·1 per cent. SnO_2.

For flotation tests two samples of natural cassiterite and one of synthetic cassiterite were prepared. Specimens of stream cassiterite (*A*) and of cassiterite from Dakota, U.S.A. (cassiterite *C*), supplied by Ward's Natural Science Establishment, were hand-picked as before, ground under water in a porcelain mortar, and deslimed by decantation. By this means material suitable for cylinder flotation tests was obtained.

A synthetic sample (cassiterite *D*) of size distribution suitable for cylinder tests was prepared by the method of Schuhmann and Brahm Prakash (1950) in which stannic oxide is recrystallized from molten borax. This sample was composed of clear colourless crystals.

(ii) *Sodium Cetyl Sulphate.*—The sample of sodium cetyl sulphate ($\mathrm{Na\overline{CS}}$ "*B*") used in this work was prepared by Evans (unpublished data). He also determined by conductivity measurements the critical concentration for micelle formation (190 mg./l.). Sutherland (1950) concluded from measurements of the surface tension of solutions of the sample that it was sensibly free from cetyl alcohol. The major impurity was sodium stearyl sulphate which constituted approximately 3 per cent. of the sample. This was noted as a slight precipitate in solutions containing more than 40 mg. sodium cetyl sulphate per litre. To remove it, stock solutions were allowed to stand until the crystals had settled, when clear supernatant solution was decanted and analysed for use.

(iii) *Water.*—The water used for adsorption tests was distilled twice from all Pyrex apparatus, alkaline permanganate being added before the second distillation.

III. Adsorption Measurements

(*a*) *Apparatus and Technique*

A quantity of cassiterite, usually about 10 mg., was added to a stud-topped Pyrex glass adsorption tube (Fig. 1) containing a hollow-glass sphere. The tube was filled completely with a solution of sodium cetyl sulphate which had been warmed to 35 °C., the temperature of the thermostat. Air was excluded so that no froth would form. Gentle agitation was provided by rotating the tubes end over end. After an adequate period in the thermostat, the tubes were placed in a centrifuge and the suspension settled by spinning for 10 minutes at 1500×g. The tubes were then placed vertically in the thermostat and aliquots taken for analysis. All tests were done in duplicate and compared with blank tests done with cassiterite absent. When there was a significant discrepancy between the concentration of reagent in the blank and the calculated composition of the original solution, the results were discarded.

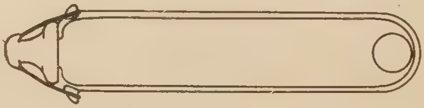


Fig. 1.—Stud-topped adsorption tube.

The aliquots of sodium cetyl sulphate solution were analysed by a variation of the methylene-blue method which has been developed by Edwards, Ewers, and Mansfield (unpublished data).

(*b*) *Results of Adsorption Measurements*

In preliminary adsorption tests with natural cassiterite (*A*), it was found that the abstraction of sodium cetyl sulphate by a suspension of mineral was due in part to adsorption on the mineral, but in part also to precipitation by ions derived from impurities in the mineral. This result suggested further that similar ions may have activated the cassiterite surface for the adsorption of sodium cetyl sulphate. These considerations led us to use synthetic cassiterite in all subsequent adsorption tests.

The amount of cetyl sulphate adsorbed was measured (i) as a function of pH, with the initial concentration chosen so that the equilibrium concentration was approximately constant, (ii) at varying concentrations for each of two constant pH values, and (iii) as a function of time. All measurements were made at $35 \pm 0{\cdot}25$ °C.

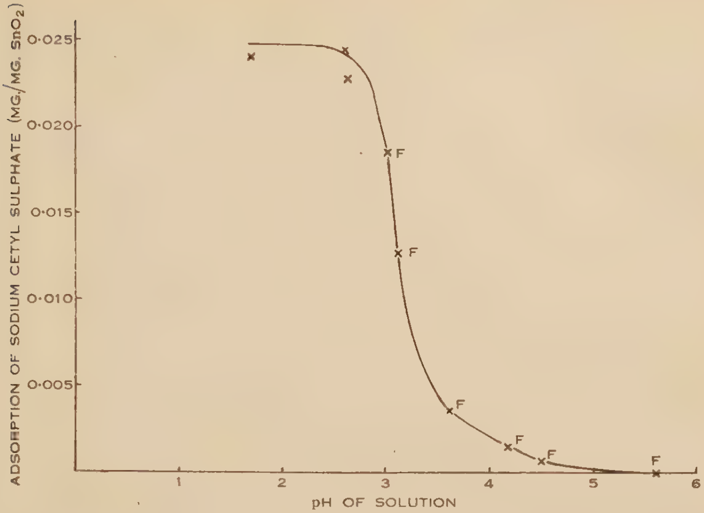


Fig. 2.—Adsorption of sodium cetyl sulphate on cassiterite from solutions of residual concentration 30±6 mg./l. and varying pH. Temperature 35 °C.

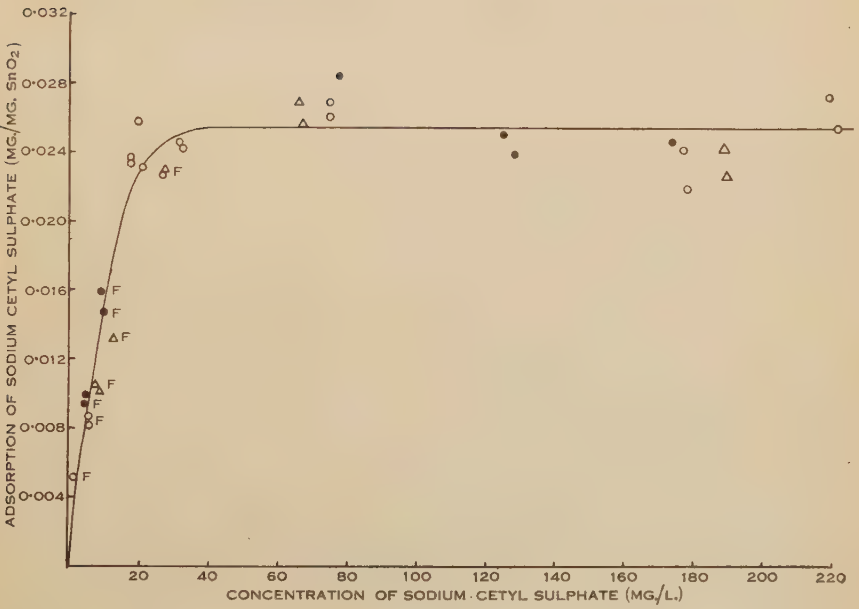


Fig. 3.—Adsorption isotherm of sodium cetyl sulphate on cassiterite from solutions of pH 2·6±0·1 and varying concentrations. Temperature 35 °C. Adsorption times: ●, 64 hours; o, 18 hours; △, 1 hour.

The results of these tests are plotted in Figures 2, 3, and 4.

It will be seen in Figure 2 that with a constant final concentration of 30 ± 6 mg. sodium cetyl sulphate per litre the adsorption is negligible until the pH is less than about 5. It increases rapidly with decreasing pH until a maximum is reached at pH 2·6. The significance of the symbol, *F*, will appear later.

The isotherms (Figs. 3, 4) at pH values of $2 \cdot 6 \pm 0 \cdot 1$ and $3 \cdot 55 \pm 0 \cdot 05$ show adsorption increasing with concentration until the same maximum is reached, after which adsorption remains sensibly constant over a wide range of concentrations.

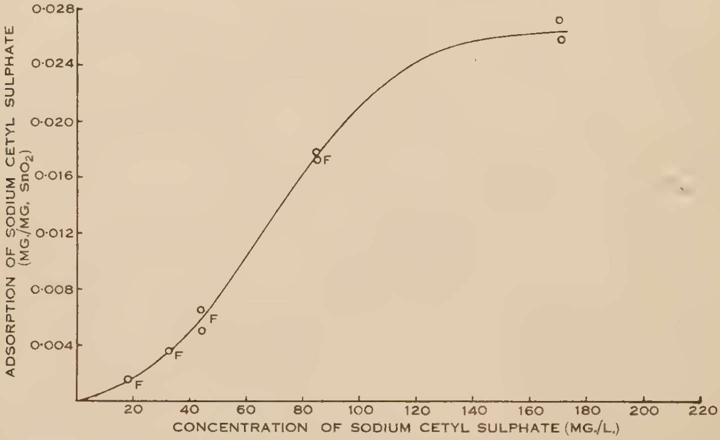


Fig. 4.—Adsorption isotherm of sodium cetyl sulphate on cassiterite from solutions of pH $3 \cdot 55 \pm 0 \cdot 05$ and varying concentrations. Temperature 35 °C. Adsorption time : 18 hours.

At concentrations of sodium cetyl sulphate below about 20 mg./l. there is a significant difference observed between the amounts adsorbed after 1 hour and 18 hours. Tests in which the agitation was continued for 64 hours indicate, however, that equilibrium is attained in something less than 18 hours.

IV. The Mechanism of Adsorption at the SnO_2–H_2O Interface

The most probable mechanism for the adsorption of sodium cetyl sulphate at the surface of cassiterite appeared to be the exchange of cetyl sulphate ion for hydroxyl ion. This is consistent with the observed increase in adsorption at low pH values. It was noted, however, that in those adsorption tests where a significant change in pH occurred during the test, an increase in hydrogen ion concentration apparently accompanied the adsorption of sodium cetyl sulphate. Thus, when the pH value of the initial solution and the blank was 4·84, the pH of the residual solution after adsorption had occurred was 4·49. To explain these results and to explore the processes occurring at the cassiterite surface a series of experiments was made in which pH changes in suspensions of the mineral were measured.

The apparatus shown in Figure 5 was made from Pyrex glass. It comprised electrode vessels *A* for the glass electrode, *B* for the saturated KCl calomel electrode, with a saturated KCl salt bridge through the three-way tap *C*. The various openings provided for the introduction of CO_2-free water or solution, dry cassiterite where necessary, from the device shown in the inset *D*, and CO_2-free air to agitate the suspension and to provide a positive pressure to prevent CO_2 and other contaminants from entering from the laboratory air. A thermometer is placed in *A*.

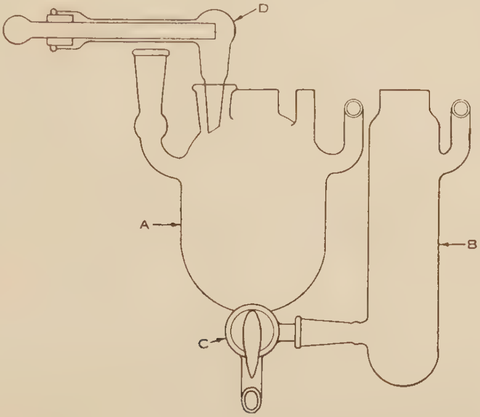


Fig. 5.—Apparatus for measuring pH changes in suspensions of cassiterite.
A, glass electrode vessel; *B*, calomel electrode vessel; *C*, three-way tap; *D*, device for adding cassiterite.

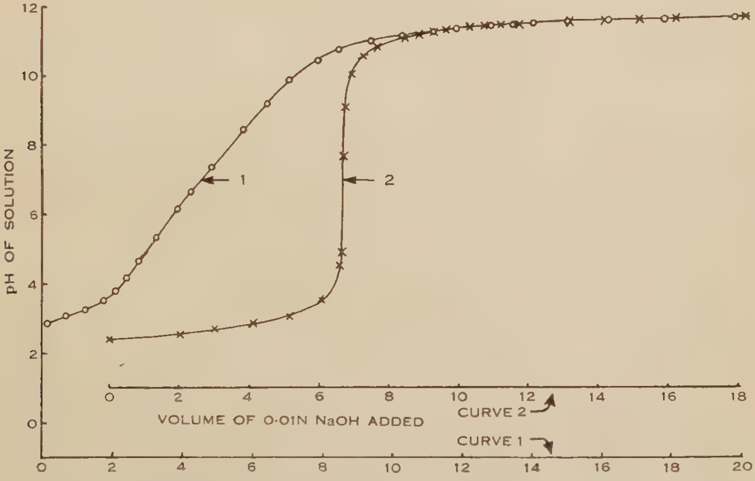


Fig. 6.—pH curves for the titration of the surface of 1·972 g. cassiterite (curve 1) and 0·0093N HCl (curve 2) with 0·01N NaOH.

In Figure 6 (curve 1) is given the titration curve obtained when N/100 NaOH was added to a suspension of cassiterite *B* in water initially acidified with HCl. It is compared with the curve (2) of a simple titration of HCl with NaOH in the absence of cassiterite. Curve 2 is displaced horizontally to make it coincide, above pH 11·4, with curve 1. The amount of acid corresponding to this displacement, added to the difference between the amounts of HCl initially present in the two titrations, gives an estimate of the acid equivalent to the cassiterite surface. Thus :

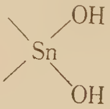
$$1\cdot972 \text{ g. } SnO_2 \equiv 8\cdot04 \text{ ml. } 0\cdot0093N \text{ HCl.}$$

In other words the surface of 1 g. cassiterite *B*, suspended in water, is capable of supplying $3\cdot8 \times 10^{-5}$ mol. hydrogen ion.

Gaudin and Rizo-Patron (1943) reported results of a similar kind obtained with powdered quartz.

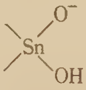
The form of the titration curve for the cassiterite surface is somewhat reminiscent of curves obtained with ion exchange resins in which the ionization of a group may be expected to influence the tendency of neighbouring groups to ionize (e.g. Bauman 1949, p. 64).

To interpret the result of this titration it is necessary to make some assumptions about the structure of the wetted cassiterite surface. If the surface be considered as being composed of amphoteric units of the type

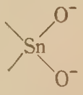


(I)

then dissociation on the alkaline side would give rise to the structure

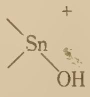


or perhaps

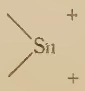


(II) (III)

On the acid side it is possible that hydroxyl ions may dissociate to give the structures corresponding to



or



(IV) (V)

From the titration curve it was not possible to determine at what pH values the various structures I to V most closely represent the actual surface.

When cassiterite was added to CO_2-free distilled water there was no significant change of pH. This was taken at first to mean that at about pH 7 the surface was substantially un-ionized, or otherwise that the structures II and IV occurred to a similar extent.

To test this theory solutions of sodium sulphate and calcium chloride were placed in the cell and the pH of the solutions measured before and after adding dry cassiterite. It was argued that, assuming the surface was neutral at pH 7, the addition of the divalent cation, Ca^{++}, ions should cause a decrease in pH, whereas if $SO_4^{=}$ ions were added liberation of OH^- would increase the pH. In fact, in both solutions, the pH decreased. These and results of similar tests are given in Table 1. In all tests 10 ml. of solution and 0·1 g. cassiterite were used.

TABLE 1

THE EFFECT ON THE PH OF ELECTROLYTE SOLUTIONS OF ADDING DRY CASSITERITE

Salt	Concentration	pH Before Addition	pH After Addition	H^+ Ions Liberated (mole/mg. SnO_2)	Area of Surface per H^+ Ion Liberated (A^2)
Na_2SO_4	50 mg. $SO_4^{=}$/l.	7·02	4·30	$5{\cdot}0\times10^{-9}$	330
Na_2SO_4	50 mg. $SO_4^{=}$/l.	6·68	4·23	$5{\cdot}9\times10^{-9}$	280
$CaCl_2$	50 mg. Ca^{++}/l.	5·70	4·20	$6{\cdot}1\times10^{-9}$	270
KCl	50 mg. K^+/l.	5·69	4·32	$4{\cdot}6\times10^{-9}$	360
$NaHCO_3$	50 mg. Na^+/l.	9·3	8·5	$3{\cdot}1\times10^{-9}$	550
$Na\overline{CS}$	40 mg. $Na\overline{CS}$/l.	7·50	4·23	$5{\cdot}9\times10^{-9}$	280

It will be noted that in all solutions, including that of sodium cetyl sulphate, there was an increase of hydrogen ion concentration, usually of the same order of magnitude.

These results suggest the following explanation. Over the range of pH values measured in the titration the cassiterite surface is ionizing to yield structure II and possibly III. The point corresponding to equal concentrations on the surface of structures II and IV is at a pH of the order 2 to 3. At about pH 7 the ionization to give II is partially complete but hydrogen counter ions are held so closely to the negative surface that the pH of the bulk of the solution is unaffected. When electrolytes are added to the suspension, a redistribution of cations occurs, the added cation partially replacing hydrogen ions from their positions in the electrical double layer, thereby reducing the pH of the suspension. This system is analogous to the Donnan membrane equilibrium, the primarily adsorbed hydroxyl ion being the counterpart of the ion to which the membrane is impervious. Similar behaviour has been noted by Steenberg (1944) who studied the adsorption of electrolytes on activated charcoal.

An attempt was made to determine an isoelectric point by studying the coagulation of the suspended cassiterite as the pH of the solution was varied. It was found that the cassiterite was coagulated when the pH lay between 0·9 and 7·6. Outside that range of pH values the cassiterite dispersed. Addition of sodium cetyl sulphate to the system raised the lower limit of coagulation.

With 30 mg. sodium cetyl sulphate per litre the cassiterite was dispersed up to pH 2·6 and with 100 mg. sodium cetyl sulphate per litre up to pH 3·5.

Sodium cetyl sulphate appears to play a dual role at the surface of the mineral. At pH 5, where the amount of cetyl sulphate adsorbed is negligible, it behaves just as do the other electrolytes tested, the sodium ion replacing hydrogen ion from the diffuse part of the double layer. At lower pH values the anion, cetyl sulphate, is adsorbed presumably by exchange with hydroxyl ions on the hydrated surface. At pH 2·5, the pH of maximum adsorption, 7×10^{-8} moles of cetyl sulphate ion is abstracted from the solution for each mg. cassiterite. This is approximately 10 times the amount of hydrogen ion released by the salt effect so that at about this pH, exchange of cetyl sulphate for hydroxyl should produce a resultant decrease in hydrogen ion concentration. Since pH measurement is not sufficiently sensitive in this range, the amount of hydroxyl released was determined by titration.

Aliquots of the supernatant solution from an adsorption test at pH 2, and of the original solution were titrated to pH 7 with 0·01N NaOH. It was found that the adsorption of $6 \cdot 0 \times 10^{-8}$ moles $Na\overline{CS}$ per mg. SnO_2 was accompanied by the release of $6 \cdot 4 \times 10^{-8}$ moles OH^- per mg. SnO_2. The residual concentration of sodium cetyl sulphate in this test was 4 mg. per litre.

It would appear, therefore, that the postulated exchange adsorption, with cetyl sulphate ion replacing adsorbed hydroxyl ion, is consistent with the observations. This mechanism is indistinguishable from the less plausible adsorption of cetyl sulphuric acid according to the neutral molecule theory of Cook and Nixon (1950).

Before considering the results of flotation tests it is convenient to summarize the results of this and the previous section in terms of the areas occupied by adsorbed molecules.

From the titration of the cassiterite surface with NaOH it was shown that 1 g. of cassiterite (10 m.^{2} of surface) was the acid equivalent of $3 \cdot 8 \times 10^{-5}$ moles of hydrogen ion. That is, there was one dissociable H^+ ion for each 44 A^2 of cassiterite surface. At about pH 7 one hydrogen ion was displaced by the addition of electrolytes from its counter ion position for every 300 A^2 of surface.

The maximum adsorption of cetyl sulphate ion, 0·025 mg. $Na\overline{CS}$ per mg. SnO_2, corresponded to one cetyl sulphate ion for each 23 A^2.

V. Flotation Tests

In this section it is intended to compare the flotation of cassiterite with adsorption occurring under similar conditions, with a view first to determining what coverage of the mineral surface by adsorbed collector is sufficient to give flotation, and secondly to finding what causes of the cessation flotation at high concentrations of collector. Some evidence of flotation was noted during adsorption tests. When the cassiterite suspension was coagulated, bubbles introduced into the tubes became completely coated with the very fine mineral. This was not a satisfactory criterion of flotation, however, since it was not possible to observe whether or not it occurred in tests where the cassiterite was well dispersed. It is interesting to note that this evidence of flotation was observed in tests marked *F* in Figures 2, 3, and 4.

Contact angle measurements were not attempted because a sufficiently large crystal of pure stannic oxide was not available and, as will be seen later, the flotation of a sample of the natural mineral depends on its composition. Furthermore, contact angle measurements on specimens of natural cassiterite were very difficult to reproduce.

The cylinder flotation test described by Wark (1938, p. 72) was modified to give a semi-quantitative measure of flotation. In the modified method, cylinders were made of Pyrex glass to the design shown in Figure 7. The short length of capillary afforded a method of measuring the amount of mineral used in the test and, after shaking, of the amount of mineral remaining unfloated. The measurement is subject to some uncertainty as the packing of the mineral before and after shaking with air is not reproducible. This uncertainty becomes more important, of course, when a small proportion of the mineral is floated, and a small difference of two relatively great lengths is taken. In practice up to 15 to 20 per cent. apparent flotation may indicate merely a difference in packing. In all tests the mineral was conditioned for one hour at 35 °C.

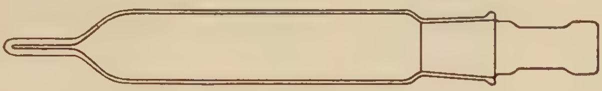


Fig. 7.—Modified apparatus for cylinder flotation tests.

Three samples of cassiterite were used. The first, the borax-crystallized synthetic cassiterite *D*, was used in tests to relate adsorption data to flotation, and the two natural cassiterites *A* and *C* were used to illustrate the variations which may occur from sample to sample.

The flotation of the three samples in cylinder tests is compared in Figure 8. A constant concentration of 20 mg. sodium cetyl sulphate per litre was used and the percentage floating measured at various pH values. It will be seen that the curves for the Dakota sample (*C*) and the synthetic sample (*D*) are very similar. The Australian sample (*A*) is markedly different. This sample contained iron ; when heated to 800 °C. it became brick red. This demonstration that the flotation of natural cassiterite differs from sample to sample may explain, at least in part, the discrepancies noted by Hergt, Rogers, and Sutherland (1946) between contact tests and cell tests. The activating effect of storing ground cassiterite under water in soda-glass bottles is shown by the fourth curve of Figure 8. The entire question of activation in this system will be the subject of a future study.

Flotation curves for a number of concentrations of sodium cetyl sulphate are shown in Figure 9. These show that, at low concentrations, flotation is favoured by low pH values, i.e. where the concentration of competing hydroxyl ion is lowest. As the concentration of collector increases so is the flotation improved at higher pH. Superimposed on this effect is the phenomenon hitherto ascribed to bubble armouring, double-layer formation, or both. Either of these processes would be aided by an increased electrolyte concentration. In solutions in which electrolyte (hydrochloric acid) has been added to give low pH values the flotation in solutions of concentrations >80 mg./l. is impaired.

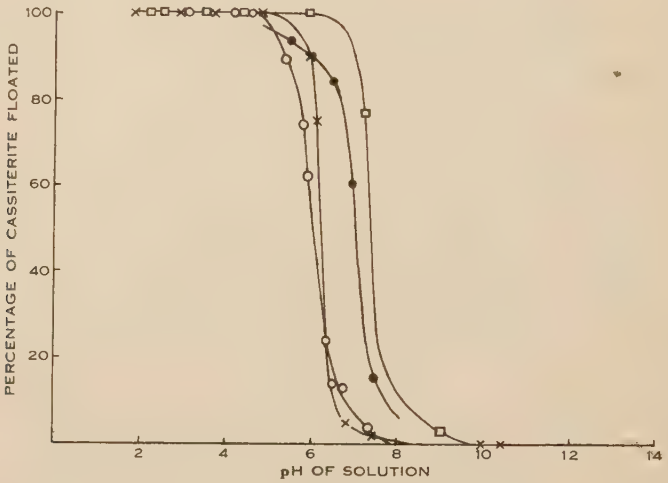


Fig. 8.—Flotation curves for cassiterite samples in a solution containing 20 mg. sodium cetyl sulphate per litre. Temperature 35 °C. □, cassiterite *A*; ×, cassiterite *C*; o, cassiterite *D*; ●, cassiterite *C*, activated by storage in soda glass.

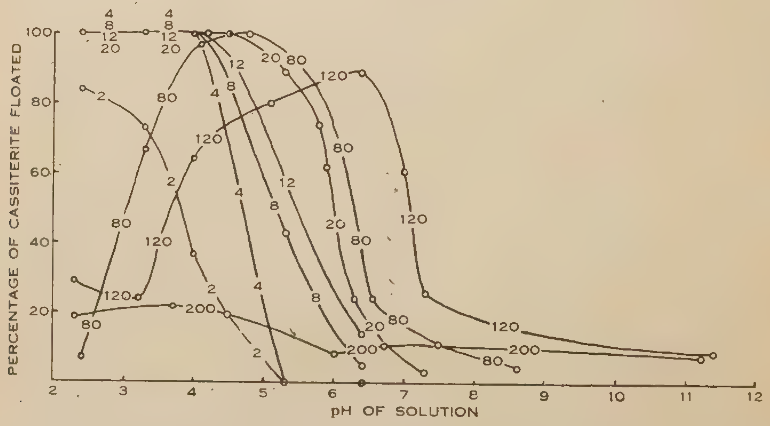


Fig. 9.—Flotation curves for cassiterite *D* in solutions of various concentrations of sodium cetyl sulphate (mg./l.). Temperature 35 °C.

Thus the curves for 80 and 120 mg./l. pass through maxima at pH 5 and 6·5 respectively where the electrolyte concentrations are relatively low. Flotation is zero or very poor in a solution of 200 mg. sodium cetyl sulphate per litre at all pH values tested. From Figure 3 it is evident that, at concentrations >80 mg./l., the amount of cetyl sulphate ion adsorbed does not differ significantly from that at, say, 20 mg./l. As excellent flotation is obtained at pH 2·5 in the latter solution, it is clear that the negligible flotation in the more concentrated solutions cannot be ascribed to either the formation of a double layer or to detergent action by the collector. Rather do the results lend support to the "bubble armouring" theory. It is known, for instance, that when electrolytes are added to solutions of soaps or similar compounds the surface activity of the compound is increased, so that in the acid solutions of sodium cetyl sulphate the surface of bubbles will be more completely covered with adsorbed ions. When sodium chloride was added in sufficient quantity to give a constant electrolyte concentration over the pH range 4 to 7, flotation of cassiterite, in a solution containing 80 mg. sodium cetyl sulphate per litre, was completely inhibited.

The curves in Figure 9 representing flotation in solutions of 4 and 20 mg. sodium cetyl sulphate per litre correspond closely to the curves obtained by Hergt, Rogers, and Sutherland (1946) for cell flotation tests with solutions containing initially 5 and 25 mg. sodium cetyl sulphate per litre respectively. That their curves for higher concentrations do not agree with ours may well be due to differences between flotation cell tests and cylinder tests with respect to bubble armouring.

The adsorption and flotation tests also provide some interesting information on the amount of collector which must be adsorbed on the mineral to permit flotation. Some of the more striking of these are set out in Table 2.

TABLE 2

THE RELATION BETWEEN AMOUNT OF SODIUM CETYL SULPHATE ADSORBED AND PERCENTAGE FLOTATION

pH	Residual Concentration NaCS (mg./l.)	Amount Na$\overline{\text{CS}}$ Adsorbed		Percentage Covered*	Percentage Floated
		mg./mg. SnO_2	Area per Ion $\overline{\text{CS}}^-$ (A^2)		
3·5	18·8	0·0016	360	6·4	100
4·5	35·7	0·0007	820	2·8	100
4·15	28·0	0·0014	410	5·6	100
5·6	36·1	Not detectable (<0·0001)	>5000	<0·4	c. 50

* The percentage of the surface covered is calculated on the assumption that each ion occupied 23 A^2 of surface.

It is interesting to note that the flotation curve for solutions of, say, 20 mg. sodium cetyl sulphate per litre is approximately parallel to the adsorption curve (Fig. 2). The displacement between them along the pH axis corresponds to about 3 pH units in a direction consistent with our conclusion that only a small fraction of the surface need be covered by collector to produce flotation.

VI. CONCLUSIONS

1. It has been confirmed that only a very small proportion of the mineral surface need be covered by collector ions to make flotation possible. No satisfactory theory has been developed, here or elsewhere, to account for this phenomenon.

2. In the particular system, cassiterite and sodium cetyl sulphate, no evidence was found for the formation of a double layer of the adsorbate. The cessation of flotation at high concentrations is attributed entirely to the armouring of the bubbles with collector adsorbed at the air-water interface. This does not necessarily constitute a rebuttal of Held's conclusions for another system, barium sulphate and sodium laurate.

3. The maximum adsorption of sodium cetyl sulphate on cassiterite corresponded to one molecule of the adsorbate for each 23 A^2 of surface.

4. We have postulated that the surface of wetted cassiterite is composed largely of amphoteric OH groups. This was consistent with the ability of the surface to act as an acid when titrated with alkali, and to adsorb cetyl sulphate ions by exchange with hydroxyl ion when the pH was sufficiently low. When the pH exceeded 4, the liberation of hydroxyl ions was masked by cation exchange in the electrical double layer from which hydrogen ion was replaced by sodium ion.

5. Samples of natural cassiterite differed quantitatively in their ability to be floated in solutions of sodium cetyl sulphate, though qualitatively they behaved similarly.

VII. ACKNOWLEDGMENTS

The authors acknowledge gratefully the assistance derived from stimulating discussions with Dr. I. W. Wark, and Dr. K. L. Sutherland and other members of the Physical Chemistry Section, Division of Industrial Chemistry, C.S.I.R.O.

VIII. REFERENCES

BAUMAN, W. C. (1949).—" Ion Exchange." (Ed. F. C. Nachod.) (Academic Press: New York.)

COOK, M. A., and NIXON, J. C. (1950).—*J. Phys. Colloid Chem.* **54**: 445.

DEAN, R. S., and AMBROSE, P. M. (1944).—U.S. Bur. Mines Bull. No. 449.

GAUDIN, A. M., and BLOECHER, F. W. (1950).—*Trans. Amer. Inst. Min.* (*Metall.*) *Engrs.* **187**: 499.

GAUDIN, A. M., and PRELLER, G. S. (1946).—*Trans. Amer. Inst. Min.* (*Metall.*) *Engrs.* **169**: 255.

GAUDIN, A. M., and RIZO-PATRON, A. (1943).—*Trans. Amer. Inst. Min.* (*Metall.*) *Engrs.* **153**: 462.

HELD, N. A., and SAMOCHWALOV, K. N. (1935).—*Kolloidzschr.* **72**: 13.

HELD, N. A., and KHAINSKY, I. A. (1936).—*Kolloidzschr.* **76**: 26.

HERGT, H. F. A., ROGERS, J., and SUTHERLAND, K. L. (1946).—*Trans. Amer. Inst. Min.* (*Metall.*) *Engrs.* **169**: 448.

NAUMOV, S. N., and MANULKIN, Z. M. (1935).—*J. Gen. Chem. U.S.S.R.*, **5**: 281.

ROGERS, J., and SUTHERLAND, K. L. (1946).—*Trans. Amer. Inst. Min.* (*Metall.*) *Engrs.* **169**: 317.

SCHUHMANN, R., and BRAHM PRAKASH (1950).—*Trans. Amer. Inst. Min. (Metall.) Engrs.* **187** : 601.

SIMARD, G. L., CHUPAK, J., and SALLEY, D. J. (1950).—*Trans. Amer. Inst. Min. (Metall.) Engrs.* **187** : 359.

STEENBERG, B. (1944).—" Adsorption and Exchange of Ions on Activated Charcoal." (Almqvist and Wiksells : Uppsala.)

STOCKTON, E. S. (1949).—*J. Sci. Instrum.* **26** : 157.

SUTHERLAND, K. L. (1950).—Thesis, University of London.

TAGGART, A. F., and ARBITER, N. (1943).—*Trans. Amer. Inst. Min. (Metall.) Engrs.* **153** : 500.

TAGGART, A. F., and HASSIALIS, M. D. (1946).—*Trans. Amer. Inst. Min. (Metall.) Engrs.* **169** : 259.

WARK, I. W. (1936).—*J. Phys. Chem.* **40** : 661.

WARK, I. W. (1938).—" Principles of Flotation." (Aust. Inst. Min. Met. : Melbourne.)

APPENDIX I

THE PREPARATION OF PURE CASSITERITE AS A FINE POWDER

By W. E. EWERS* and K. F. LORKING*

A sample of pure stannic oxide was prepared by burning tetraethyl-tin, atomized in a stream of air, in a hydrogen flame.

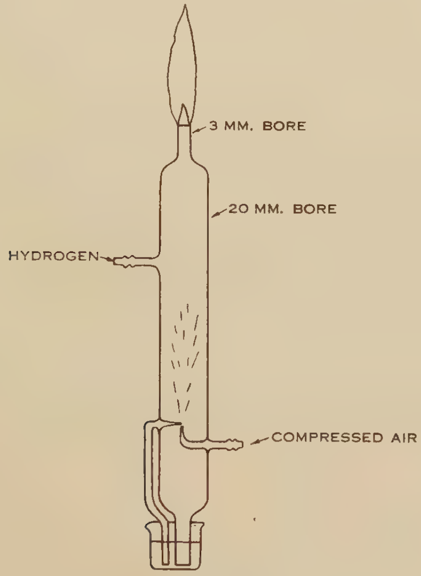


Fig. 10.—Burner for producing stannic oxide smoke.

To prepare tetraethyl-tin a method similar to that described by Naumov and Manulkin (1935) was used. Stannic chloride (b.p. 114–115 °C. at 760 mm.) was dissolved in a mixture of diethyl ether and low-boiling light petroleum and was treated with ethyl magnesium bromide. After refluxing for 1½ hours, the excess Grignard reagent was hydrolysed with dilute acetic acid. The ether layer was separated, dried over calcium chloride, and fractionally distilled. The yield of tetraethyl-tin as a colourless liquid boiling between 180 and 181° C. (760 mm.) was 80 per cent. calculated on the amount of stannic chloride used.

---

* Division of Industrial Chemistry, C.S.I.R.O., Melbourne.

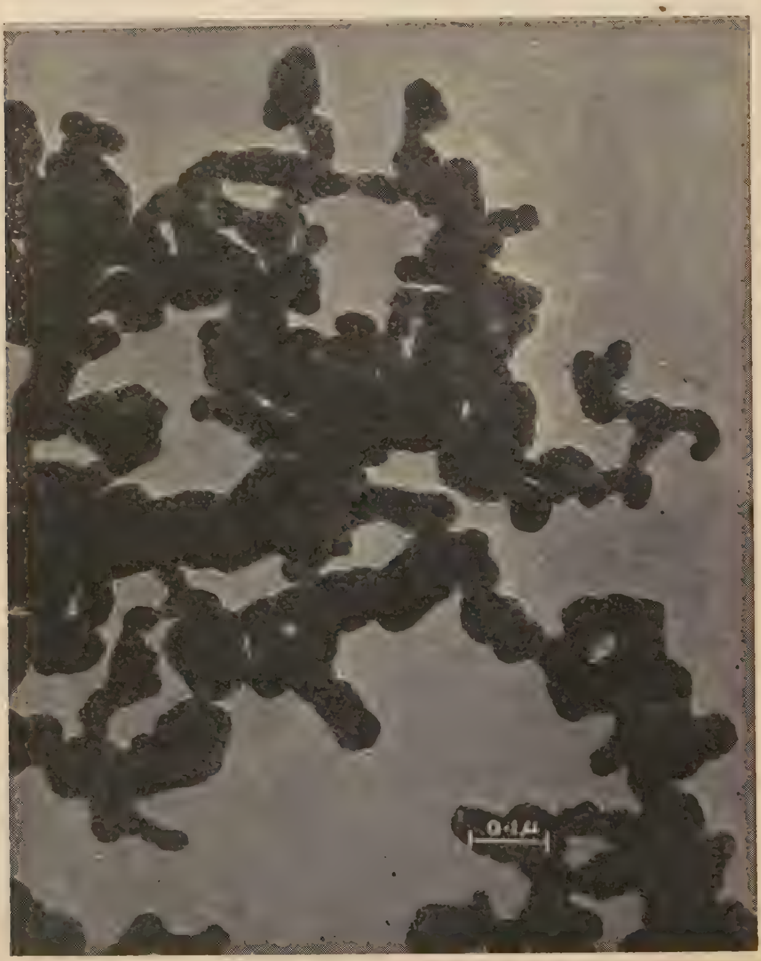


Fig. 11.—Electron micrograph of stannic oxide smoke. ×100,000.

In burning tetraethyl-tin it was necessary to avoid exposing the liquid or vapour to a hot surface. Burners, in which the compound was heated or vaporized in a tube, rapidly became blocked as the decomposition products built up on the heated surface. For the same reason burners in which liquid or vapour were led through a jet directly into a flame could not be used. Wick burners failed because the gas velocity in the flame was insufficient to remove the solid oxide which fell back on the wick and rapidly snuffed the flame.

The burner chosen was made of glass and was adapted from the design of Stockton (1919). A diagram of the burner and collecting apparatus is shown in

Figure 10. It is essentially a Bunsen burner in which tetraethyl-tin, atomized by a stream of air, is premixed with hydrogen before burning. Droplets of tetraethyl-tin too large to reach the flame fall back into the beaker.

The combustion products, including the stannic oxide smoke, were drawn by gentle suction into an all glass system comprising a funnel, a water condenser, and a condensate trap. The stannic oxide was collected in two fractions. About one-third of it accumulated in the funnel above the burner as a fine white powder. The part collected in the condenser and flask was slightly grey but when calcined subsequently at 800 °C. it became almost white, retaining a very slight yellowish colour.

The product was examined by X-ray diffraction by A. McL. Mathieson and positively identified as cassiterite. Chemical analysis showed that both fractions contained $100 \cdot 0 \pm 0 \cdot 1$ per cent. SnO_2. Electron micrographs, taken by J. L. Farrant, with a magnification ×100,000, showed regular crystals ranging in size from 250 to 1000 Å (Fig. 11). The surface area was determined by nitrogen adsorption using the B.E.T. method. Assuming the area, S, of the adsorbed nitrogen molecule to be $16 \cdot 2$ A^2, the surface area was 10 $m.^2/g.$ Based on the electron micrograph and making no allowance for the degree to which the particles appear to be sintered together, an estimate of 12 $m.^2/g.$ was obtained for the surface area.

THE THERMAL DECOMPOSITION OF THORIUM OXALATE

By Roberta Beckett* and M. E. Winfield*

[*Manuscript received July* 13, 1951]

Summary

The thermal decomposition of hydrated thorium oxalate was followed by determining weight loss, change in electrical conductivity, and chemical composition of products. At temperatures below 200 °C. the weight-loss curve indicated the existence of a tri- and a dihydrate and possibly a pentahydrate of thorium oxalate as stages in the dehydration. At higher temperatures dehydration was accompanied by loss of CO and CO_2.

Almost all the $(C_2O_4)^=$ decomposed sharply over a small temperature range near 300 °C. Although some carbonate was formed, it did not exceed a concentration of 0·075 mole per mole of total thorium.

According to chemical analysis the residue after heating at 350 °C. was ThO_2 with a trace of carbonate and water. X-ray analysis, however, showed it to be a mixture of ThO_2 and a second substance whose identity is unknown.

In the electron microscope the hydrated oxalate and the thoria produced from it were little different in appearance. Both appeared as relic crystals which were square prisms about 1 μ across and 0·2 μ thick, with a characteristic central aperture which exposed a large area of surface.

I. Introduction

It has long been known that the oxide prepared by heating thorium oxalate is not identical with that obtained from salts such as the nitrate or sulphate (Kohlschütter and Frey 1916). In order to prepare the most active oxide for the catalytic dehydration of 2,3-butanediol to 1,3-butadiene, Winfield (1945) found that it was necessary to heat thorium oxalate, rather than other salts, to about 350 °C. Traces of nitrate in the oxalate were sufficient to modify the course of the decomposition so that the product was an unsatisfactory catalyst. When it was used at temperatures of 450 °C. or more, the active thoria gradually became less effective.

Before the mechanism of the abovementioned catalysis can be learned, it is clearly necessary to know the structure of the so-called thoria on whose surface the reaction proceeds. Besides examining the structure of the catalytically active material by X-ray diffraction and electron microscopy, we have studied some of the physical and chemical changes which take place when hydrated thorium oxalate is gradually raised in temperature. The earliest of these changes influences the structure of the final product.

II. Experimental

Thorium oxalate was prepared in the manner previously described (Winfield 1945). After drying at 40 °C. and atmospheric pressure it was allowed to come

---

* Division of Industrial Chemistry, C.S.I.R.O., Melbourne.

to equilibrium with the atmosphere at 20 °C. Gold was evaporated on to the two ends of an oxalate pellet 2 mm. long and 5 mm. in diameter. Its electrical conductivity was measured while the temperature was raised to 515 °C. over the course of 90 hours.

Another measure of the progress of decomposition with temperature was obtained by heating an oxalate sample in an automatic dehydration apparatus, which recorded equilibrium weights as a function of temperature.

For determination of the chemical composition of the mixture of compounds present at a given temperature, a sample of $Th(C_2O_4)_2$, $5\cdot48\ H_2O$ contained between two watch-glasses was weighed before and after heating for 2 hours at the required temperature, then analysed. The watch-glasses, which were 1·5 cm. in diameter, were ground to fit at the edges so that atmospheric water was excluded from the hygroscopic products when the temperature was lowered. With this technique the solid is enveloped in its gaseous products during thermal decomposition. The results therefore do not correspond exactly with those obtained by the first two methods which have been described.

Thorium carbonate estimations were made by measuring the volume of CO_2 produced when 2N sulphuric acid solution was added to each sample. Thorium oxalate was determined by permanganate titration. The remainder of the thorium* was assumed to be ThO_2. Water was estimated from the difference between the total weight and the sum of the weights of the three thorium compounds.

For electron microscopy, oxalate samples which had been heated at temperatures from 40 to 650 °C. were dispersed in amyl alcohol. Some of the low-temperature samples were pumped slowly in a vacuum desiccator before mounting, in order to minimize damage to the crystals during rapid pumping in the electron microscope.

III. Results

The variation of conductivity with temperature (Fig. 1 (*a*)) showed little resemblance to the curve (Fig. 1 (*c*)) given by Hahn and Senftner (1934), who measured thoron production during thermal decomposition of an oxalate sample containing radiothorium. Recently we have been able to check the conductivity curve by the weight-loss method (Fig. 1 (*b*)). Below the point *B* in this curve there are a number of spurious equilibria, as in the decomposition of calcium oxalate (Zawadzki and Bretsznajder 1938). Their presence was clearly demonstrated by bursts of decomposition each time the temperature was raised during the conductivity measurements. When curves 1 (*a*) and 1 (*b*) are replotted in comparable units, for example percentage change, points *A* and *B* are found to refer to the same transition. We cannot explain with certainty this break at *A* or *B* in the oxalate decomposition. Chemical analyses indicate that it corresponds with the point of maximum thorium carbonate concentration. Up to *A* and *B* the decomposition of $(C_2O_4)^{-}$ may have proceeded largely through the carbonate.

---

* Total thorium was known from the weight and composition of the sample before heating.

From the weight-loss curve it was calculated that the product at 180 °C. has the composition $Th(C_2O_4)_2,2H_2O$, which was confirmed by chemical analysis. Between 180 and 290 °C. the shape of the curve suggested that this region did not correspond simply to loss of a further mole of water per mole of thorium. Analysis of the products subsequently showed that in this range CO together with a little CO_2 was lost as well as water, as a preliminary to the extensive decomposition of $(C_2O_4)^=$ which commenced at 295 °C. and was practically isothermal, except for a displacement of about 10 °C. at point *B*.

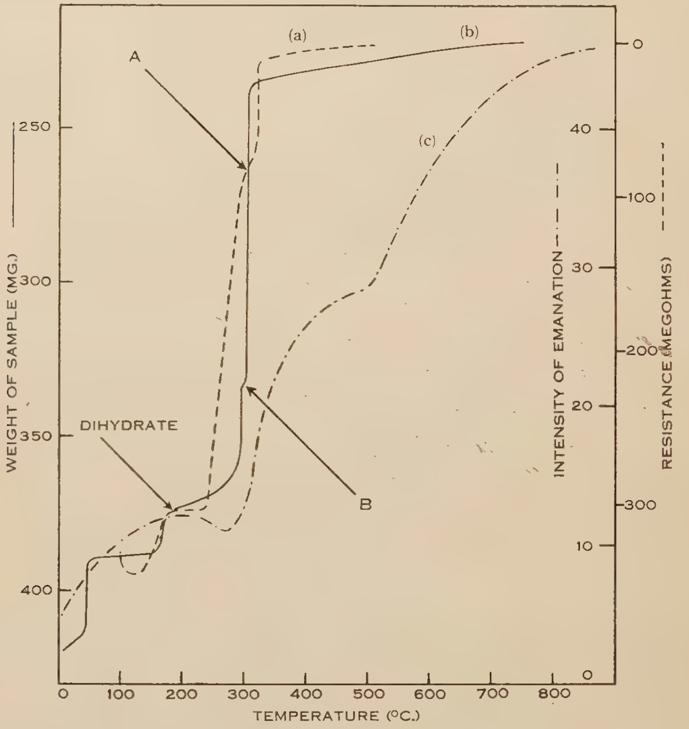


Fig. 1.—Weight-loss and conductivity curves showing the dehydration and decomposition of thorium oxalate as a function of temperature. The result obtained by Hahn and Senftner using the emanation method is reproduced for comparison.

Above 320 °C. the gradual loss in weight, which continued to more than 600 °C., was found to be due not only to removal of water but also to break down of some carbonate which persisted to high temperatures (Fig. 2). Although Figure 3 suggests that some oxalate resisted decomposition (for example, 0·3 per cent. at 367 °C.), this result may be explained by formation in a side reaction of a trace of carbon, which also reacts with the permanganate used in oxalate estimation (Gunther and Rehaag 1938).

No peroxide could be detected in the decomposing oxalate by means of titanium sulphate reagent (Gunther and Rehaag 1938). It was therefore assumed that the peroxydioxalate suggested as an intermediate by these authors was absent from our products, and that our CO_2 determination could be taken as a measure of the amount of thorium carbonate in the mixture.

As shown in Figure 2 the carbonate concentration during oxalate decomposition reached a maximum* of only 0·075 mole per mole of total thorium. The decomposition temperatures of thorium oxalate and thorium carbonate are therefore close together, which accounts for the frequent failure to observe the formation of carbonate as an intermediate in the conversion of $Th(C_2O_4)_2$ to ThO_2. Indeed our results suggest that some of the ThO_2 may be produced directly from oxalate. Figure 3 shows that ThO_2 may appear at temperatures as low as 200 °C.

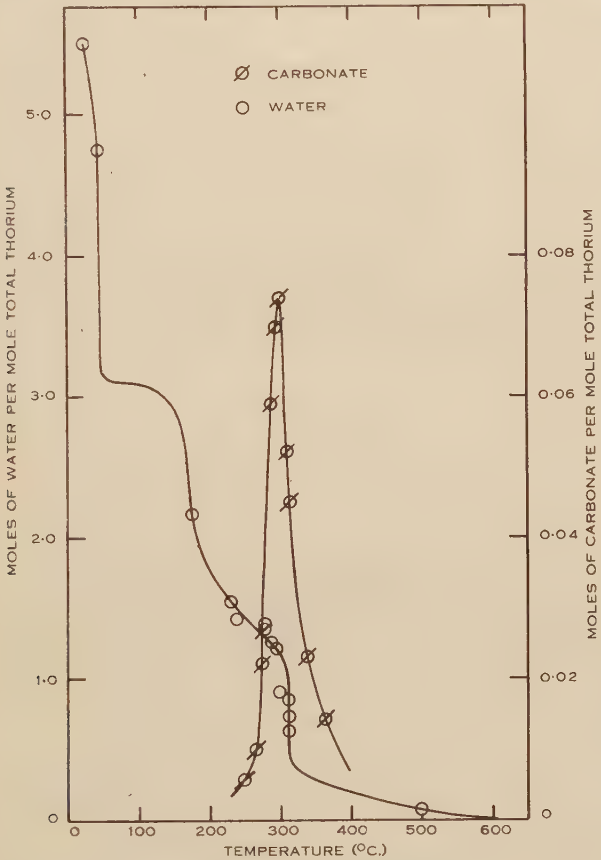


Fig. 2.—Carbonate concentration together with an approximate representation of the dehydration which occurs during thermal decomposition of thorium oxalate.

---

* The maximum at 300 °C. in Figure 2 probably corresponds with a point at 295 °C. in the weight-loss curve in Figure 1. Temperatures in the two kinds of experiment are not strictly comparable because of the different gaseous environment.

From the chemical and physical evidence presented above, we can conclude that the catalyst which was especially active in the dehydration of 2,3-butanediol to butadiene (Winfield 1945) consisted essentially of about 98 per cent. ThO_2, together with small amounts of water and thorium carbonate. The X-ray evidence, however, shows that the active surface cannot be assumed to have the interatomic spacings generally accepted for crystalline thorium oxide.

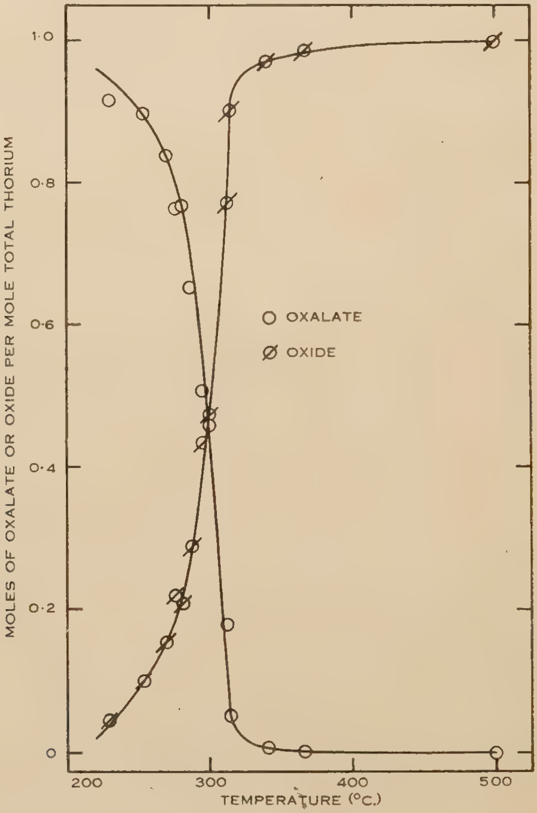


Fig. 3.—Oxalate and oxide concentrations in the product obtained by heating thorium oxalate.

The dehydration curve in Figure 2 was prepared by interpolating, with the aid of information in Figure 1, between the points determined by chemical analysis. Figures 1 and 2 indicate the existence of a dihydrate, as found for lanthanum and yttrium oxalates by Wylie (1947), as well as tri- and possibly pentahydrates. When the oxalate decomposition is near completion, about one-half mole of water per mole of thoria still remains. It is possible that the loss of catalytic activity above 450 °C. (Winfield 1950) is due to structural changes which accompany loss of this water.

X-ray powder patterns show that the 350 °C. product contains besides thorium oxide (lattice parameter a_0=5·58 Å) a second substance, also cubic, with a unit cell 2·3 times as large. We have not been able to identify the latter constituent, which is not oxalate as stated by Winfield (1950) nor carbon, and can suggest only that it may be a hydrate of ThO_2. On heating at 500 °C. the powder patterns for both substances become sharper, presumably due to crystallite growth. At the same time the intensity of the unknown lines decreases. Even after heating at 800 °C., which further sharpens both sets of lines, those corresponding to $a_0 = 13 \cdot 0$ Å are not entirely extinguished. We therefore have no proof that the loss in catalytic activity at 500 °C. is related to the decrease in concentration of the unknown constituent rather than the increase in size of the crystallites.

At all compositions from $Th(C_2O_4)_2,5 \cdot 48\ H_2O$ to ThO_2 the crystals or their relics have a characteristic central aperture (Plate 1) whose presence in the highly hydrated material was checked with a high-power optical microscope. The large holes which are present in the particles of the catalyst are therefore not formed by shrinkage during loss of water, CO, and CO_2. It is tentatively suggested in explanation that the approximately cubic crystals of hydrated thorium oxalate are formed by recrystallization, during the aging process, around a core of the initial amorphous precipitate that is obtained even in hot acid solution. Subsequent handling and drying lead to loss of the structurally-weak core.

It is apparent in Plate 1 why thoria obtained from the oxalate has a large surface area. The surface within the central hole is not smooth like the exterior, but is broken by projecting laminae so thin as to be transparent to electrons. Both by shadowing with palladium and by splitting the backing film to turn a crystal on edge, the thoria relic crystals can be shown to be one-fifth as thick as they are long, whereas crystals of hydrated oxalate are roughly cubic. This, and the occurrence of detached fragments which are plate-like, indicates that the relic crystals are laminated in a direction parallel to the large sides.

That a large proportion of the total surface is provided by the laminae which project into the central aperture is supported by a surface-area calculation, based on a crystal of average size. If the walls of the aperture are smooth, the total geometric area is 2 sq. m./g. compared with 25 sq. m./g. as determined by the B.E.T. method (Emmett 1941).

In many micrographs the spongy nature of the individual lamellae is clearly visible. One thoria specimen showed a regular cellular structure, each cell being a cube about 100 Å across. The cells may be relics of the mosaic structure in the original crystal of hydrated oxalate. The observed sponginess is in accord with the apparent density in benzene, which is 5·5 g./cc. compared with a true density of 10·0.

From the above descriptions it will be seen that thoria prepared from the oxalate is unusually interesting catalytic material, in that it combines the chemical composition of catalyst with the physical structure of catalyst support.

After pumping at 0·001 mm. pressure at 20 °C., samples of $Th(C_2O_4)_2,5 \cdot 48\ H_2O$ had the composition $Th(C_2O_4)_2,1 \cdot 7\ H_2O$, while at 100 °C. the water content was reduced to 0·67 mole per mole of thorium. Thus it must be assumed that in the electron microscope none of the oxalates have a water content exceeding

1 mole per mole of thorium, because of the low pressure and the heating effect of the electron beam. From a comparison of all micrographs taken, it was concluded that a crystal changes little in outward appearance as its composition is varied in stages from $Th(C_2O_4)_2,H_2O$ to ThO_2. A minor change which was detected in the fine structure may be due to contraction in wall thickness of the sponge, of which the crystal is composed.

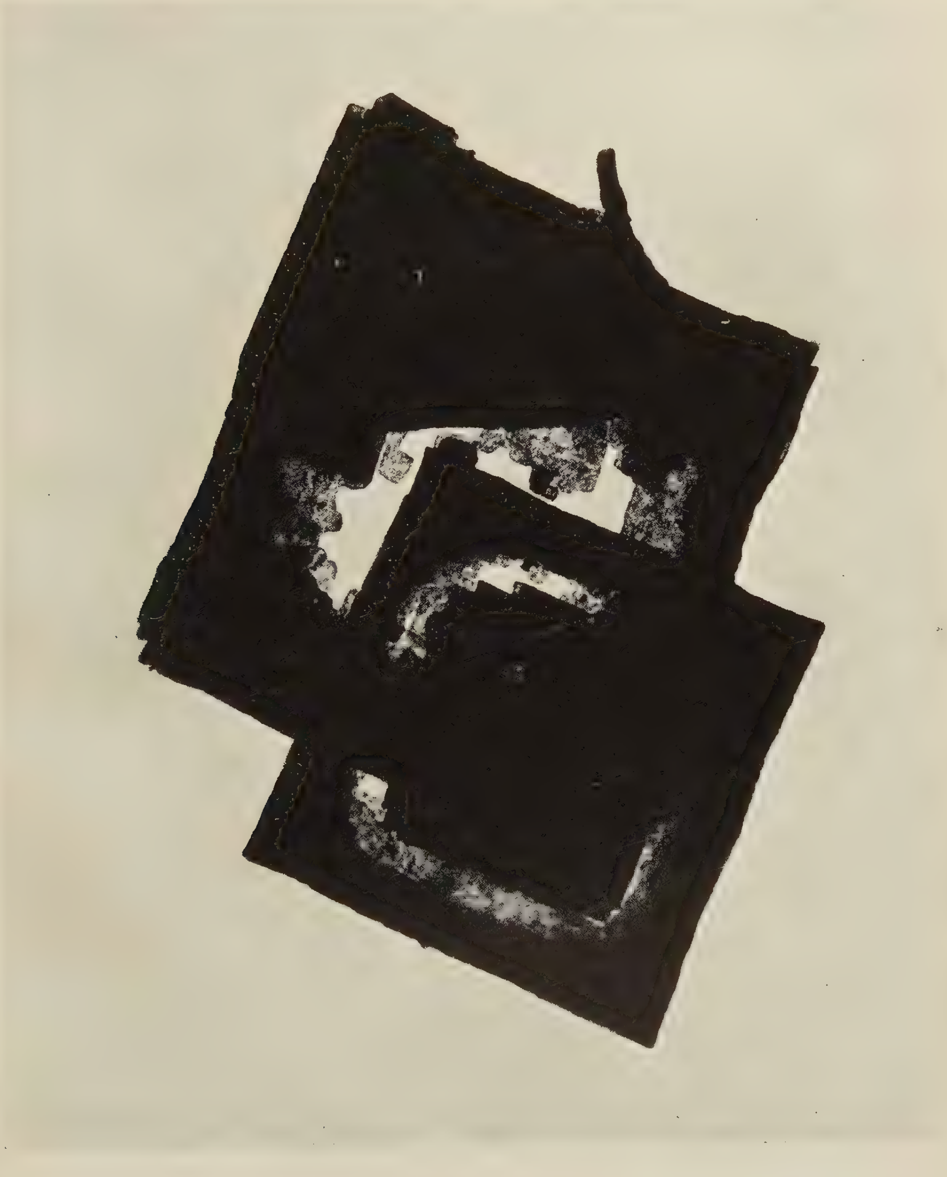
IV. ACKNOWLEDGMENTS

We wish to thank several members of the Division of Industrial Chemistry, C.S.I.R.O., for assistance; in particular Dr. J. Cowley for X-ray patterns, Mr. A. Gaskin for the thermal weight-loss determination, and Mr. G. Edwards for CO_2 estimations.

V. REFERENCES

EMMETT, P. H. (1941).—*Proc. Amer. Soc. Test. Mater.* **41**: 95.
GUNTHER, P. L., and REHAAG, H. (1938).—*Ber. dtsch. chem. Ges.* **71B**: 1771-7.
HAHN, O., and SENFTNER, V. (1934).—*Z. phys. Chem.* A **170**: 191-211.
KOHLSCHÜTTER, V., and FREY, A. (1916).—*Z. Elektrochem.* **22**: 145-61.
WINFIELD, M. E. (1945).—*J. Coun. Sci. Industr. Res. Aust.* **18**: 412-23.
WINFIELD, M. E. (1950).—*Aust. J. Sci. Res.* A **3**: 290-305.
WYLIE, A. W. (1947).—*J. Chem. Soc.* **1947**: 1687-92.
ZAWADZKI, J., and BRETSZNAJDER, S. (1938).—*Trans. Faraday Soc.* **34**: 951-9.

THERMAL DECOMPOSITION OF THORIUM OXALATE



Electron micrograph of thoria produced at 650 °C. ×55,200 ; not shadowed.

COORDINATION COMPLEXES OF INDIUM AND 2,2′,2″-TERPYRIDYL AND OF INDIUM AND 5-NITRO-1,10-PHENANTHROLINE

By G. J. Sutton*

[*Manuscript received May* 31, 1951]

Summary

Indium ions react with 2,2′,2″-terpyridyl in dilute alcohol or in aqueous solution forming bisterpyridyl indium salts, of which the chloride, bromide, iodide, and thiocyanate are insoluble. As thiocyanate, one part of indium per 10^5 may be detected. Metals such as zinc, tin, and iron, with which indium is usually associated, interfere with the reaction.

With 5-nitro-1,10-phenanthroline, indium forms insoluble tris-(5-nitro-1,10-phenanthroline) indium salts as chloride, bromide, iodide, and thiocyanate. The reactions are of little value since the least soluble salt, the iodide, is 0·04 per cent. soluble.

I. Bisterpyridyl Indium Salts

The ready formation of chelate compounds of indium with 2,2′-dipyridyl and 1,10-phenanthroline (Sutton 1949) suggested the possibility of tridentate coordination with 2,2′,2″-terpyridyl in aqueous solution. The unavailability of the reagent necessitated the preparation of a small amount of it, using a less efficient but necessary modification by combining the methods described by Wibaut and Willink (1931) and Morgan and Burstall (1932). This was accomplished by the bromination of pyridine in the presence of iron filings and catalysts in a steel autoclave. Bisterpyridyl ferrous bromide sesquihydrate was obtained in the products, and a small amount of the base subsequently isolated. The other products were neglected. Morgan and Burstall (1937) have described the formation of monotridentate complexes of the base, [MX tripy] and [MX tripy]X, where M=Cu and Ag; and [D tripy X]X, where D=Zn, Cd, Hg, Pd. Bistridentate complexes [M 2tripy]X_2, [M 2tripy]X_3, $n H_2O$, where M=Fe, Ru, Os, Co, Ni, and Cr were also studied. With indium ions, terpyridyl produces white crystalline precipitates of the chloride, bromide, and thiocyanate. The three salts may be crystallized in the anhydrous state from acetone. An insoluble yellow iodide is also formed, which is colourless in acetone indicating the probable absence of a diacido complex.

On analysis, all four compounds were shown to consist of two molecules of base to one of indium salt. In contrast to the uncoordinated indium halides, the bisterpyridyl complexes are not deliquescent. They are destroyed by strong acids or by alkalis.

---

* Chemistry Department, N.S.W. University of Technology, Sydney.

II. TRIS-(5-NITRO-1,10-PHENANTHROLINE) INDIUM SALTS

In neutral or slightly acid solutions, 5-nitro-1,10-phenanthroline coordinates with indium ions forming tris-(5-nitro-1,10-phenanthroline) indium salts, of which the chloride, bromide, and thiocyanate are white and the least soluble the iodide is yellow in colour. All four salts dissolve readily in the lower alcohols, 1,4-dioxane, and acetone, giving colourless solutions.

III. EXPERIMENTAL

(a) *2,2′,2″-Terpyridyl*

Pyridine (60 ml.), bromine (30 ml.), fine iron filings (15 g.), and a mixture of pumice powder, activated carbon, and cuprous chloride in equal proportions (5 g.) were heated in a small cylindrical steel autoclave of about 500 ml. capacity, internal diameter 5 cm. and walls 1 cm. thick. After 12 hours at 200 °C., the autoclave and contents were allowed to cool spontaneously to room temperature. The contents were transferred to a beaker together with water (100 ml.) acidified with hydrochloric acid, and heated for 20 minutes at 90 °C. The mixture was then made alkaline with sodium hydroxide solution and steam distilled. Excess alkali was added and the oily layer which separated was extracted with light petroleum (3×25 ml.). The light petroleum was distilled off, and the small brownish residue steam distilled. The distillate yielded an almost white compound which melted at 87 °C. (0·12 g.).

(b) *Bisterpyridyl Indium Salts*

Freshly sublimed indium chloride (0·036 g.) was dissolved in water (4 ml.). Terpyridyl (0·064 g.) was dissolved in ethanol (8 ml.) and diluted with water (to 40 ml.). Portion of the latter solution (10 ml.) was treated with indium chloride solution (1 ml.), shaken, and allowed to stand (1 hour). A very fine white silky precipitate settled out. The residual solutions were divided into three equal portions respectively, and to the terpyridyl solutions were added the calculated proportions of potassium bromide (0·015 g.), potassium iodide (0·020 g.), and potassium thiocyanate (0·012 g.) each in water (1 ml.) respectively. The indium chloride solutions were then treated with the terpyridyl solutions containing the appropriate potassium salt, and allowed to stand. The bromide was precipitated as very fine white crystals, and the thiocyanate as a heavy white precipitate. The respective suspensions were centrifuged (10 minutes at 3000 r.p.m.), decanted, washed with water (2 ml.), recentrifuged, decanted, partially dried by a hot-air blower, and recrystallized from acetone. Owing to the small amounts of material, it was desirable to have the compounds in the anhydrous state, for analytical purposes. Non-deliquescent crystals of each salt were obtainable from acetone, methanol, ethanol, and 1,4-dioxane. Owing to the difficulty in obtaining terpyridyl, only very small amounts of material were available for analysis, accounting for the analytical data of Table 1, the results of which are only in approximate agreement.

TABLE 1

ANALYTICAL DATA—BISTERPYRIDYL INDIUM SALTS

Substance	Yield (%)	Indium		Anion		Decomp. (° C.)
		Found (%)	Calculated (%)	Found (%)	Calculated (%)	
In(terpy)$_2$Cl$_3$	90	16·8	16·7	16·8	15·5	320
In(terpy)$_2$Br$_3$	98	15·0	14·0	28·8	29·2	290
In(terpy)$_2$I$_3$	99	11·2	12·0	37·3	39·6	310
In(terpy)$_2$(CNS)$_3$ ..	97	14·8	15·1	24·0	23·0	150

Indium chloride (0·0083 g.) together with potassium thiocyanate (0·1 g.) in water (100 ml.) and 1% terpyridyl (1 ml.) gave a white opalescence in a few seconds indicating a reaction of 5 parts per 10^5 of indium.

(c) *Tris-(5-nitro-1,10-phenanthroline) Indium Salts*

Indium chloride (0·09 g.) was dissolved in water (20 ml.) and divided into four equal portions. 5-Nitro-1,10-phenanthroline (0·270 g.) was dissolved in ethanol (12 ml.) and divided into four equal portions. By mixing a portion of each solution fine pale bistre crystals of tris-(5-nitro-1,10-phenanthroline) indium chloride were obtained. Potassium bromide (0·036 g.), potassium iodide (0·05 g.), and potassium thiocyanate (0·03 g.) were added to the residual proportions of indium chloride respectively, and each allowed to react with a portion of the 5-nitro-1,10-phenanthroline solution. Small pale bistre crystals of the bromide and thiocyanate were obtained, whilst the iodide formed canary-yellow clusters. The suspensions were separated, washed, and dried as in the case of the terpyridyl complexes, and crystallized from methanol. The complexes are not deliquescent and may be crystallized from ethanol. An inspection of the filtrates after centrifuging indicated a decrease in solubility from chloride, bromide, thiocyanate, to iodide.

Table 2

ANALYTICAL DATA—TRIS-(5-NITRO-1,10-PHENANTHROLINE) INDIUM SALTS

Substance	Indium		Anion		Decomp. (° C.)
	Found (%)	Calculated (%)	Found (%)	Calculated (%)	
$In(nitphen)_3Cl_3$	12·7	12·7	11·9	11·8	207
$In(nitphen)_3Br_3$	11·0	11·1	23·3	23·3	181
$In(nitphen)_3I_3$	9·7	9·8	32·6	32·5	179
$In(nitphen)_3(CNS)_3$ ..	12·0	11·9	18·0	18·1	162

In Table 2 the indium was estimated gravimetrically as In_2O_3 (Scott 1947) using semi-micro equipment.

IV. References

Morgan, G. T., and Burstall, F. H. (1932).—*J. Chem. Soc.* **135** : 20–30.

Morgan, G. T., and Burstall, F. H. (1937).—*J. Chem. Soc.* **1937** : 1649-55.

Scott, W. W. (1947).—"Standard Methods of Chemical Analysis." 5th Ed. Vol. 1. pp. 446-47. (Van Nostrand Co. Inc. : New York.)

Sutton, G. J. (1949).—*Aust. Chem. Inst. J. & Proc.* **16** : 115-9.

Wibaut, J. P., and Willink, H. D. T. (1931).—*Rec. Trav. Chim. Pays-Bas* **56** : 287-90.

STUDIES IN THE CHEMISTRY OF SOME THALLIUM III COMPLEXES

By G. J. Sutton*

[*Manuscript received June* 18, 1951]

Summary

Thallium III ions react with alkali alkylxanthates, forming insoluble yellow xanthates $(ROCS_2)_3Tl$. In accordance with the properties of tervalent metal xanthates as inner complexes, they are soluble in diethyl ether, and in common organic solvents.

Aqueous or dilute alcoholic solutions of 1,10-phenanthroline and 2,2′-dipyridyl react with thallium III ions forming white insoluble chlorides, bromides, and thiocyanates as Tl phenan X_3 and Tl dipy X_3 respectively. The iodides are from orange-red to orange-yellow in colour and are very insoluble in water. The empirical formulae of the latter are Tl phenan$_2$ I_3 and Tl dipy$_2$ I_3 respectively; and tests indicate that the sensitivity of the iodide reactions enables 1 part of thallium in 10^6 parts of solution to be detected in the presence of iron, in the ferric state.

I. Thallium III Alkylxanthates

The formation of some Thallium I and Thallium III xanthates was observed by Atsuki and Takata (1939) who mentioned their instability, and a detailed study was made by Anderson (1950). In accordance with the observations that xanthates of tervalent metals are soluble in diethyl ether, it was proposed to investigate the general reactions of the thallium III alkylxanthates. In slightly acid solutions the latter may be prepared as insoluble yellow compounds by double decomposition with the corresponding alkali xanthates. They are appreciably soluble in common organic solvents, and are readily decomposed by strong acids.

When heated, they decompose at temperatures between 150 and 160 °C. In general, they resemble the indium xanthates described by the author (1950) and a similar tris-chelated structure is suggested. The compounds of the lower alkyl members are readily crystallizable from ether, whilst the higher homologues form somewhat gummy compounds from ethereal solutions.

II. Thallium III–2,2′–Dipyridyl Reactions

In its lower valency, a bisdipyridyl thallium I nitrate (Tl dipy$_2$ $NO_3,2\frac{1}{2}H_2O$) was observed by Pfeiffer and Christeleit (1938). In aqueous solution or in dilute alcohol, thallium III chloride is coordinated by 2,2′-dipyridyl with the formation of an insoluble complex having an empirical formula Tl dipy Cl_3. Similar bromide and thiocyanate complexes may be obtained. They are soluble in the lower alcohols and in acetone, and are slightly soluble in hot water.

---

* Chemistry Department, N.S.W. University of Technology, Sydney.

Two possible structures are (*a*) dimeric bridge complexes ($Tl_2\,dipy_2\,X_6$) and (*b*) $Tl\,dipy\,X_3$, which is in agreement with the tendency for thallium to be stabilized by an electron quartet as in $Tl(CH_3)_2Cl_3$ etc.

In the presence of an iodide, however, an orange-yellow to orange-red precipitate may be obtained, which is almost insoluble in hot water and insoluble in cold. The compound which analyses as $Tl\,dipy_2\,I_3$ may be crystallized from acetone, in which it is very soluble. In dilute acetone, in the presence of potassium iodide, the complex gives no reaction with starch, which indicates a $(Tl^{III}\,dipy\,I_2)I$ structure, in contrast with thallium III iodide (or thallium I periodide) in which an I_3^- ion is indicated. Hieber and Sonnekalb (1928) ascribed a similar diacido structure to ethylenediamine thallium III iodide ; although this latter compound could have a dimeric bridge structure. Treatment of $Tl\,dipy_2\,I_3$ with $AgNO_3$ and even with AgCl in dilute acetone by prolonged refluxing results in total iodide precipitation, and colourless filtrates. Both sodium thiosulphate and sodium sulphite, however, reduce the complex to $Tl^{I}\,dipy_2\,I$.

Variation in colour was found to be due to particle size and not, as was first thought, to *cis-trans* isomerism. Grinding to fine particle size gives a consistent orange-yellow colour. The $Tl\,dipy_2\,I_3$ may be obtained directly from the other $Tl\,dipy\,X_3$ salts or from $Tl\,en\,I_3$ in alcohol. At 120 °C. partial decomposition into $Tl^{I}\,dipy_2\,I$ takes place ; the reaction being effected more rapidly at about 190 °C. Treatment of the residue after prolonged heating at 120 °C. with starch and potassium iodide indicates tri-iodide. Thus $(Tl^{I}\,dipy_2)I_3$ forms as an intermediate product. The $Tl^{I}\,dipy_2\,I$ formed by reduction with sulphite or thiosulphate, or by pyrolysis at 190 °C., is almost insoluble in acetone, but is slightly soluble in dilute alcohol. Treatment of $Tl^{I}\,dipy_2\,I$ with potassium periodide solution results firstly in the formation of greyish $(Tl^{I}\,dipy_2)I_3$; which is transformed to the orange $(Tl^{III}\,dipy_2\,I_2)I$ in the presence of excess potassium periodide.

III. Thallium III-1,10-Phenanthroline Reactions

Repetition of the foregoing using 1,10-phenanthroline in lieu of 2,2′-dipyridyl, results in a similar series such as $Tl\,phenan\,X_3$, where X=chloride, bromide, or thiocyanate ; and $Tl\,phenan_2\,I_3$ in the presence of iodides. Attempts to form complexes of thallium I and 1,10-phenanthroline by direct addition give negative results. However, $Tl^{I}\,phenan_2\,I$ may be obtained by pyrolysis of $Tl^{III}\,phenan_2\,I_3$ below 200 °C.

IV. Experimental

Thallium acetate was oxidized to thallium III hydroxide and the chloride $TlCl_3,4H_2O$ obtained from the latter. A solution containing $TlCl_3,4H_2O$ (0·0192 g./ml.) was prepared.

(*a*) *Thallium III Alkylxanthates*

The following experimental procedure applies to the preparation of the thallium III xanthates. Thallium III chloride solution (20 ml.) was treated with an excess (2 g.) of the appropriate sodium alkylxanthate in water (50 ml.) with stirring.

The heavy yellow amorphous precipitates, which filtered slowly, were washed twice with water (5 ml.) and then with ethanol (5 ml.). Treatment with diethyl ether (50 ml.) resulted in pale yellow solutions, from which yellow rhombic crystals were obtained, in the case of the methyl,

ethyl, and *iso*propyl derivatives. The *n*-butyl and secondary amylxanthates yielded apparently amorphous products. The xanthates were dried to constant weight over calcium chloride in a vacuum desiccator (3 days). They were found to be soluble in pyridine, benzene, toluene, acetone, 1,4-dioxane, and *cyclo*hexane. They showed only slight solubility in carbon tetrachloride, chloroform, methanol, and ethanol. The yields were up to 90% in the case of lower alkyl members, whilst only about 60% yields were obtained with the butyl and amyl derivatives. Their analyses and temperatures of decomposition are summarized in Table 1.

TABLE 1

ANALYTICAL DATA—THALLIUM III ALKYLXANTHATES

Formula	Thallium Found (%)	Thallium Calculated (%)	Melting Point (Decomp.) (°C.)
$(CH_3OCS_2)_3Tl$	39·2	38·9	150
$(C_2H_5OCS_2)_3Tl$	36·2	36·0	159
$(iso\text{-}C_3H_7OCS_2)_3Tl$	33·6	33·5	163
$(n\text{-}C_4H_9OCS_2)_3Tl$	31·8	31·3	158
$(sec.\text{-}C_5H_{11}OCS_2)_3Tl$ ..	28·9	29·4	148

(*b*) *Thallium III-2,2′-Dipyridyl Complexes*

Dipyridyl (0·156 g.) was dissolved in ethanol (2 ml.) and the solution added to excess thallium III chloride solution (25 ml.) with stirring. The white precipitate which settled readily was separated by centrifuging (2500 r.p.m.), decanted, washed, recentrifuged, and decanted. The precipitate which appeared to be crystalline lost negligible weight (0·3%) at 100 °C. after drying by hot-air blower. From acetone, white radiating prisms were obtained. Repetition of the above procedure adding the calculated proportion of potassium bromide or potassium thiocyanate in each case to the reagent gave similar white precipitates. Separation by centrifuging and crystallization from acetone yielded white prisms with Tl dipy Br_3, and white silky spangles with Tl dipy $(CNS)_3$. The chloride, bromide, and thiocyanate were found to be soluble in acetone, methanol, ethanol, 1,4-dioxane, sparingly soluble in cold water, and slightly soluble in hot water. The solubility in water was found to decrease progressively from chloride, bromide, to thiocyanate. The yields were almost quantitative in the presence of excess anion. Repetition of the above, using excess dipyridyl (0·156 g.) and thallium III chloride solution (6·67 ml.) resulted in 0·154 g. substance, as Tl dipy Cl_3, the theoretical yield being 0·156 g. Using a threefold excess of dipyridyl with quantities of potassium bromide and potassium thiocyanate proportional to Thallium III chloride respectively, gave yields which corresponded to monodipyridyl complexes: bromide complex 0·202 g. (Tl dipy Br_3 requires 0·200 g.) and thiocyanate complex 0·178 g. (Tl dipy $(CNS)_3$ requires 0·178 g.). Addition of excess of the appropriate thallium III salt to the clear solutions after separation of the above by centrifuging gave yields proportional to the excess base respectively. Thus, complexes Tl dipy X_3 only were obtainable in the case of chloride, bromide, and thiocyanate from aqueous solutions with variable proportions of the reagents.

Thallium III chloride solution (20 ml. = 0·384 g.) was treated with excess dipyridyl (0·468 g.) in ethanol (3 ml.) containing potassium iodide (0·5 g.) in water (2 ml.). The orange precipitate which formed was removed by centrifuging, washed with water (10 ml.), and separated as before; then dissolved in acetone (10 ml.) and allowed to crystallize on a water-bath. Orange-red prisms were obtained, the yield (0·896 g.) corresponded to Tl $dipy_2$ I_3 (0·897 g.). The presence of excess thallium salt as iodide was avoided to prevent the formation of TlI_3 which is difficult to separate from the complex. Apart from its great solubility in acetone, the substance was found to be sparingly soluble in ethanol, benzene, toluene, 1,4-dioxane, and carbon tetrachloride. The greater solubility of this complex and the corresponding salt with 1,10-phenanthroline in organic solvents is possibly due to the high proportion of organic addenda in the cation.

The iodide complex (Tl dipy$_2$ I$_3$, 0·0758 g.) lightened in colour after heating at 190 °C. for 1 hour giving a greyish residue (0·0540 g.) which corresponded to the loss of 2 iodide atoms leaving Tl dipy$_2$ I (which requires 0·0543 g.).

Pyrolysis of the dipyridyl complexes gave the following results: Tl dipy Cl$_3$, m.p. 254 °C. (decomp. 300 °C.); Tl dipy Br$_3$, m.p. 220 °C. (decomp. 300 °C.). The iodide complex darkened at 120 °C. due to iodine formation and lightened at 190 °C., with the evolution of iodine. The residue (Tl^{I} dipy$_2$ I) did not decompose below 340 °C. Analyses of the complexes are given in Table 2, the thallium being estimated as thallium I chromate gravimetrically.

TABLE 2

ANALYTICAL DATA—THALLIUM III-2,2′-DIPYRIDYL COMPLEXES

Substance	Thallium		Anion	
	Found (%)	Calculated (%)	Found (%)	Calculated (%)
Tl dipy$_2$ I$_3$	22·5	22·7	42·6	42·5
Tl dipy Cl$_3$	43·8	43·7	22·9	22·9
Tl dipy Br$_3$	34·4	34·1	40·5	40·0
Tl dipy (CNS)$_3$	38·2	38·2	32·4	32·5
Tl dipy$_2$ I	31·7	31·8	19·9	19·7

(c) *Thallium III-1,10-Phenanthroline Complexes*

The aforementioned experimental procedure for thallium III-dipyridyl complexes was repeated in the preparation of the thallium III-1,10-phenthroline complexes. Phenanthroline (0·180 g.) in ethanol (2 ml.) was allowed to react with excess thallium III chloride solution (25 ml. = 0·48 g.). Separation by centrifuging and decantation resulted in short white needles of the chloride complex 0·479 g. from acetone. The bromide complex and thiocyanate complex were prepared similarly by the addition of potassium bromide (0·45 g.) and potassium thiocyanate (0·40 g.) respectively, to the thallium III chloride solutions. White prisms (0·601 g.) of the bromide and white silky acicular crystals (0·547 g.) of the thiocyanate were obtained. Tests showed a progressive decrease in solubility in water from chloride, bromide, to thiocyanate. Similarly 1,10-phenanthroline (0·360 g.) in ethanol (3 ml.) was added to potassium iodide (0·50 g.) in water (2 ml.) and the mixture was added to thallium III chloride solution (20 ml. = 0·384 g.). Subsequent separation and recrystallization from acetone yielded the orange iodide complex (0·930 g.) as long acicular crystals. Pyrolysis of the latter (0·0641 g.) for 4 hours at 190 °C. gave a yellowish residue (0·0465 g.) which recrystallized from acetone, in which it is sparingly soluble, as pale yellow octahedra. The loss in weight (27·4%) corresponded very closely to 2 iodine atoms (26·9%) with the formation of bisdipyridyl thallium I iodide. Treatment of thallium I acetate (0·0263 g.) in water (20 ml.), together with potassium iodide (0·0166 g.) in water (1 ml.), yielded a pale yellow precipitate (0·0335 g.) which was found to be thallium I iodide only (requires 0·0331 g.). Thus the complex Tl^{I} phenan$_2$ I was unobtainable by direct combination. The results are presented in Table 3.

Temperatures of decomposition were: chloride 280 °C., bromide 238 °C., and thiocyanate 195 °C.

Thallium III chloride solution (1 ml.), containing 0·0102 g. thallium was diluted progressively and tested with 2 ml. of a solution of 1,10-phenanthroline (0·0036 g.) and potassium iodide

TABLE 3

ANALYTICAL DATA—THALLIUM III-1,10-PHENANTHROLINE COMPLEXES

Substance	Thallium		Anion	
	Found (%)	Calculated (%)	Found (%)	Calculated (%)
Tl phenan Cl_3	41·7	41·6	21·5	21·7
Tl phenan Br_3	33·0	32·8	38·4	38·5
Tl phenan $(CNS)_3$	36·2	36·6	30·8	31·2
Tl $phenan_2$ I_3	21·6	21·6	40·2	40·3

(0·0050 g.) in 50% alcohol (10 ml.). A distinct yellow coloration developed in a dilution of the thallium solution to 10,000 ml.; indicating the detection of 1 part of the thallium in 10^6 parts of water.

V. ACKNOWLEDGMENT

The analyses of the thallium III alkylxanthates were carried out by Mr. J. R. A. Anderson, Chemistry Department, N.S.W. University of Technology.

VI. REFERENCES

ANDERSON, J. R. A. (1950).—*Aust. Chem. Inst. J. & Proc.* **17**: 120-2.

ATSUKI, R., and TAKATA, T. (1939).—*Cellulose Ind. Tokyo* **15**: 321-7; Engl. Summ. 67-9. (*Chem. Abstr.* **34**: 5053 (1940).)

HIEBER, W., and SONNEKALB, F. (1928).—*Ber. dtsch. chem. Ges.* **61** B: 555-7.

PFEIFFER, P., and CHRISTELEIT, W. (1938).—*Z. Anorg. Chem.* **239**: 133-44.

SUTTON, G. J. (1950).—*Aust. Chem. Inst. J. & Proc.* **17**: 249-53.

CORRIGENDUM

VOLUME 4, NUMBER 3

Page 279, Table 3: The following two sets of values were omitted at the beginning of the table:

Energy (keV.)	a_1/a_0	a_2/a_0	a_3/a_0	a_4/a_0	a_5/a_0	a_6/a_0
300	0·057	—0·051	0·026	0·021	—0·032	0·011
400	0·147	—0·002	0·027	—0·017	—0·006	0·012

The next two sets of values which appear in the table as printed are results at an energy of 400 keV.

INDEX

PAGE

Acridone Alkaloids, The Synthesis of the 423
Alcohols, Long-Chain *n*-Primary, The Dielectric Properties of 365
Alcohols, Long-Chain Secondary, The Dielectric Properties of .. 213
Alexander, A. E.—*See* Cumper, C. W. N. 372
Alkaloids, Acridone, The Synthesis of the 423
Allen, C. W.—*See* Christiansen, W. N., *et al.*.. 51
Angular Distribution of α-Particles from the $Li^7(p,\alpha)He^4$ Reaction, Further Measurements of the 284
Angular Distribution of α-Particles and Protons from Li^6+D^2 268
Angyal, S. J., and Warburton, W. K.—
  Sulphonamides. II. Structure and Tautomerism of Sulphapyridine, Sulphathiazole, and Sulphanilylbenzamidine 93
Anticyclones, Steady-State, The Structure of 329
Atmosphere and Ocean, A Survey of the Stress between 315
Atoms, Highly Ionized, Collision Processes Involving .. 437
Azoxybenzenes, Steric Effects in 185

Bayliss, N. S., and Rivière, J. C.—
  Conjugated Compounds. III. Spectra and Ionization in Polyenes and Polymethines by a Finite Difference Method 344
Beckett, Roberta J., and Winfield, M. E.—
  The Thermal Decomposition of Thorium Oxalate .. 644

PAGE

Benzene-*n*-heptane, The System 198
Bolton, J. G., and Westfold, K. C.—
  Galactic Radiation at Radio Frequencies. IV. The Distribution of Radio Stars in the Galaxy 476
Bottomley, W., and White, D. E.—
  The Chemistry of Western Australian Plants.
    IV. *Duboisia hopwoodii* .. 107
    V. Vellein from *Velleia discophora* .. 112
Breyer, B., Gutmann, F., and Hacobian, S.—
  Polarography with Alternating Currents. III. A.C. Polarography of Bismuth, Antimony, and Tin with Special Reference to the Presence of Halogen Ions .. 595
Breyer, B., and Hacobian, S.—
  Polarography with Alternating Currents.
    IV. A.C. Polarography of Uranium, Iron, Nickel, and Manganese 604
    V. A.C. Polarography of Some Common Anions 610
Brown, I., and Ewald, A. H.—
  Liquid-Vapour Equilibria. II. The System Benzene-*n*-heptane 198
Bruce, C. F.—
  Transmission-Like Multiple Beam Reflection Interferometry 117

Carman, E. H.—*See* Kannuluik, W. G. 305
Caro, D. E., Parry, J. K., and Rathgeber, H. D.—
  The Momentum Spectrum and Positive Excess of Cosmic Ray Mesons at Sea-Level .. 16

PAGE

Cassiterite, The Adsorption of Sodium Cetyl Sulphate on .. 627
Cellulose, The Crystalline Structure of 412
Christiansen, W. N., Hindman, J. V., Little, A. G., Payne-Scott, Ruby, Yabsley, D. E., and Allen, C. W.—
  Radio Observations of Two Large Solar Disturbances .. 51
Chronograph, A High Speed .. 526
Collision Processes Involving Highly Ionized Atoms .. 437
Conductivity of Air, Thermal, The Temperature Dependence of the 305
Conjugated Compounds .. 344
Corrigenda .. 229, 436, 658
Cosmic Ray Mesons at Sea-Level, The Momentum Spectrum and Positive Excess of 16
Cosmic Rays, The Time of Flight of 526
Crystal Structure of Potassium Hexafluoroplatinate (IV) .. 406
Crystalline Structure of Cellulose 412
Cumper, C. W. N., and Alexander, A. E.—
  Some Electrophoretic Studies of Proteins in the Adsorbed State 372
Cygnus, Constellation of, Observations of the Source of Radio-Frequency Radiation in the 158

Dielectric Absorption in Solid Aliphatic Methyl Esters .. 616
Dielectric Properties of a Series of *dl-n*-Eicosanols 359
Dielectric Properties of Crystalline Long-Chain *n*-Primary Alcohols 365
Dielectric Properties of Symmetrical Long-Chain Secondary Alcohols 213
Diffusion Measurements by the Lamm Scale Method .. 84
Doe, A. G.—*See* Tasny-Tschiassny, L. 231
Dryden, J. S., and Welsh, H. K.—
  Dielectric Absorption in Solid Aliphatic Methyl Esters .. 616

PAGE

Duboisia hopwoodii 107
Dunbar, D. N. F., and Hirst, F.—
  Angular Distribution of α-Particles and Protons from Li^6+D^2 268

Edwards, G. R., and Ewers, W. E.—
  The Adsorption of Sodium Cetyl Sulphate on Cassiterite 627
dl-n-Eicosanols 359
Electrons, The Straggling of .. 450
Electrophoretic Studies of Proteins in the Adsorbed State 372
Emission, Solar Radio-Frequency from Localized Regions at Very High Temperatures 131
Emulsions, Nuclear, The Surface Loss Correction for .. 296
Esters, Methyl, Dielectric Absorption in 616
Ewald, A. H.—*See* Brown, I... 198
Ewers, W. E.—*See* Edwards, G. R. 627

Falk, J. E., and Willis, J. B.—
  The Infra-Red Spectra of Porphyrins and their Iron Complexes 579
Foster, D. H., and Wardrop, A. B.—
  The Crystalline Structure of Cellulose as Revealed by Acid Hydrolysis 412

Galactic Radiation .. 258, 459
Galvanometer Amplifier, An Analysis of the 560
Gillam, D. G.—
  Aqueous Diffusion Measurements by the Lamm Scale Method and their Correlation with Membrane Diffusion Measurements .. 84
Gillis, R. G.—
  Small-Scale Preparation of Methyl Red Indicator .. 435

PAGE

Gore, P. H., and Hughes, G. K.—
Studies of the Wallach Transformation. II. Steric Effects in Some Symmetrically Disubstituted Azoxybenzenes 185
Gutmann, F.—*See* Breyer, B., and Hacobian, S. 595

Hacobian, S.—
See Breyer, B. .. 604, 610
See Breyer, B., and Gutmann, F. 595
Halogen Ions, A.C. Polarography of 595
Harper, A. F. A.—*See* Wylie, R. G. 560
Heat Flow Problem, Solution of a 12
Hexafluoroplatinate, Potassium 406
Hill, E. R.—
Collision Processes Involving Highly Ionized Atoms .. 437
Hindman, J. V. — *See* Christiansen, W. N., *et al.* .. 51
Hines, K. C.—
The Straggling of Electrons below the Critical Energy .. 450
Hirst, F.—*See* Dunbar, D. N. F. 268
Hirst, F., and Uebergang, R.—
Further Measurements of the Angular Distribution of α-Particles from the $\text{Li}^7(p,\alpha)\text{He}^4$ Reaction .. 284
Hughes, G. K.—*See* Gore, P. H. 185
Hughes, G. K., and Ritchie, E.—
Experiments on the Synthesis of the Acridone Alkaloids .. 423
Hydrogenation and Adsorption of Gases on Transition Metals 385

Impedance Matrices of *n*-Phase Partially Symmetrical Machines 544
Indium and 2,2′,2″-Terpyridyl, Coordination Complexes of .. 651
Infra-Red Absorption Bands, The Influence of Slit-Width on the Shape and Intensity of 172
Infra-Red Spectra of Porphyrins and their Iron Complexes 579
Interferometry, Transmission-Like Multiple Beam .. 117

PAGE

Ionization and Spectra in Polyenes and Polymethines .. 344
Ions, Halogen, A.C. Polarography of 595

James, R. W.—
The Structure of Steady-State Anticyclones .. 329

Kannuluik, W. G., and Carman, E. H.—
The Temperature Dependence of the Thermal Conductivity of Air 305

Lamm Scale Method, Measurement of Aqueous Diffusion by the 84
Lawrance, R.—
The Photovoltaic Effect in Natural Lead Sulphide .. 569
Lead Sulphide, the Photovoltaic Effect in 569
Lichtblau, H.—*See* Muirhead, E. G. 432
Liquid-Vapour Equilibria .. 198
Little, A. G.—
See Christiansen, W. N., *et al.* 51
See Payne-Scott, Ruby .. 508
Little, A. G., and Payne-Scott, Ruby—
The Position and Movement on the Solar Disk of Sources of Radiation at a Frequency of 97 Mc/s. I. Equipment 489

Martin, S. L.—
The Surface Loss Correction for Nuclear Emulsions .. 296
Meakins, R. J., and Mulley, Joan W.—
The Dielectric Properties of Crystalline Long-Chain *n*-Primary Alcohols at Low Frequencies 365
Meakins, R. J., and Sack, R. A.—
The Dielectric Properties of Symmetrical Long-Chain Secondary Alcohols in the Solid State 213
Meakins, R. J., and Welsh, H. K.—
The Dielectric Properties of a Series of *dl*-*n*-Eicosanols .. 359

PAGE

Mellor, D. P., and Stephenson, N. C.—
  The Crystal Structure of Potassium Hexafluoroplatinate (IV) 406
Mesons, Cosmic Ray, The Momentum Spectrum and Positive Excess of 16
Metals, Transition, Adsorption and Hydrogenation of Gases on 385
Methyl Esters, Dielectric Absorption in 616
Methyl Red Indicator, Small-Scale Preparation of .. 435
Microscope Objectives, Reflecting, The Design of 1
Mills, B. Y., and Thomas, A. B.—
  Observations of the Source of Radio-Frequency Radiation in the Constellation of Cygnus 158
Minnett, H. C.—*See* Piddington, J. H. 131, 459
Moore, A. J. W., and Tegart, W. J. McG.—
  Rupture of Oxide Films during Repeated Sliding 181
Muirhead, E. G., and Lichtblau, H.—
  A Circuit for the Operation of a Synchrotron in Conjunction with a Wilson Cloud Chamber 432
Mulley, Joan W.—*See* Meakins, R. J. 365

5-Nitro-1, 10-phenanthroline and Indium Coordination Complexes 651
Noise Storms 508

Ocean and Atmosphere, A Survey of the Stress between 315
Officer, V. C.—
  A High Speed Chronograph and its Application to the Time of Flight of Cosmic Rays 526
Oxide Films, Rupture of, during Repeated Sliding 181

Parry, J. K.—*See* Caro, D. E., and Rathgeber, H. D. .. 16

PAGE

α-Particles from the Li$^7(p,\alpha)$He^4 Reaction 284
α-Particles and Protons from Li^6+D^2 268
Payne-Scott, Ruby—
  See Christiansen, W. N., *et al.* 51
  See Little, A. G. 489
Payne-Scott, Ruby, and Little, A. G.—
  The Position and Movement on the Solar Disk of Sources of Radiation at a Frequency of 97 Mc/s. II. Noise Storms 508
Photovoltaic Effect in Natural Lead Sulphide 569
Piddington, J. H., and Minnett, H. C.—
  Observations of Galactic Radiation at Frequencies of 1210 and 3000 Mc/s. .. 459
  Solar Radio - Frequency Emission from Localized Regions at Very High Temperatures 131
Polarography with Alternating Currents .. 595, 604, 610
Polymethines and Polyenes, Spectra and Ionization in .. 344
Polynomial Equations, The Solution of, with the Aid of the Electrolytic Tank .. 231
Porphyrins and their Iron Complexes, The Infra-Red Spectra of 579
Potassium Hexafluoroplatinate (IV), The Crystal Structure of 406
Priestley, C. H. B.—
  A Survey of the Stress between the Ocean and Atmosphere 315
Proteins in the Adsorbed State, Electrophoretic Studies of .. 372
Protons and α-Particles from Li^6+D^2 268

Radiation at a Frequency of 97 Mc/s., Sources of 489, 508
Radiation, Enhanced 36
Radiation, Galactic 258, 459, 476
Radiation in the Constellation of Cygnus 158

PAGE

Radiation, Solar, at Metre Wavelengths 36
Radio Observations of Two Large Solar Disturbances .. 51
Radio Stars in the Galaxy, The Distribution of 476
Radok, J. R. M.—
  Solution of a Heat Flow Problem 12
Rathgeber, H. D.—*See* Caro, D. E., and Parry, J. K. .. 16
Ritchie, E.—*See* Hughes, G. K. 423
Rivière, J. C.—*See* Bayliss, N. S. 344

Sack, R. A.—*See* Meakins, R. J. 213
Shain, C. A.—
  Galactic Radiation at 18·3 Mc/s. 258
Shen, D. W. C.—
  Operational Impedance Matrices of *n*-Phase Partially Symmetrical Machines .. 544
Sodium Cetyl Sulphate, The Adsorption of 627
Solar Disk, The Position of Sources of Radiation on the 489, 508
Solar Disturbances, Radio Observations of 51
Solar Radiation at Metre Wavelengths 36
Solar Radio-Frequency Emission from Localized Regions 131
Spectra and Ionization in Polyenes and Polymethines 344
Spectra, Infra-Red, of Porphyrins and their Iron Complexes 579
Spectrum of High-Intensity Solar Radiation at Metre Wavelengths 36
Spectrum, Momentum, of Cosmic Ray Mesons at Sea-Level 16
Steel, W. H.—
  The Design of Reflecting Microscope Objectives .. 1
Stephenson, N. C.—*See* Mellor, D. P. 406
Sulphonamides, The Structure and Tautomerism of .. 93

PAGE

Sutton, G. J.—
  Coordination Complexes of Indium and 2,2′,2″-Terpyridyl and of Indium and 5-Nitro-1, 10-phenanthroline 651
  Studies in the Chemistry of Some Thallium III Complexes 654
Synchrotron in Conjunction with a Wilson Cloud Chamber, Circuit for 432

Tasny-Tschiassny, L., and Doe, A. G.—
  The Solution of Polynomial Equations with the Aid of the Electrolytic Tank .. 231
Tautomerism and Structure of Sulphonamides 93
Tegart, W. J. McG.—*See* Moore, A. J. W. 181
2,2′,2″-Terpyridyl and Indium, Coordination Complexes of.. 651
Thallium III Complexes, Studies in the Chemistry of.. 654
Thomas, A. B.—*See* Mills, B. Y. 158
Thorium Oxalate, The Thermal Decomposition of 644
Titanium, Thermoelectric Properties of 62
Transition Metals, Adsorption and Hydrogenation of Gases on 385

Uebergang, R.—*See* Hirst, F... 284

Vellein from *Velleia discophora* 112

Wallach Transformation, Studies of the 185
Warburton, W. K.—*See* Angyal, S. J. 93
Wardrop, A. B.—*See* Foster, D. H. 412
Welsh, H. K.—
  See Dryden, J. S. 616
  See Meakins, R. J. 359
Westfold, K. C.—*See* Bolton, J. G. 476
White, D. E.—*See* Bottomley, W. 107, 112

PAGE

Wild, J. P.—
Observations of the Spectrum of High-Intensity Solar Radiation at Metre Wavelengths. IV. Enhanced Radiation 36

Willis, J. B.—
The Influence of Slit-Width on the Shape and Intensity of Infra-Red Absorption Bands 172
See Falk, J. E. 579

Winfield, M. E.—
Adsorption and Hydrogenation of Gases on Transition Metals 385
See Beckett, Roberta J. .. 644

PAGE

Worner, H. W.—
Thermoelectric Properties of Titanium with Special Reference to the Allotropic Transformation 62

Wylie, R. G., and Harper, A. F. A.—
An Analysis of the Galvanometer Amplifier and its Response to Alternating Electromotive Forces and Mechanical Vibrations .. 560

Yabsley, D. E. — *See* Christiansen, W. N., *et al.* 51

DATE DUE

DEC 17 '84 PER
RET'D DEC 17 1984

PERIODICALS MUST BE RETURNED
TO PERIODICALS DESK ONLY

DEMCO 38-297